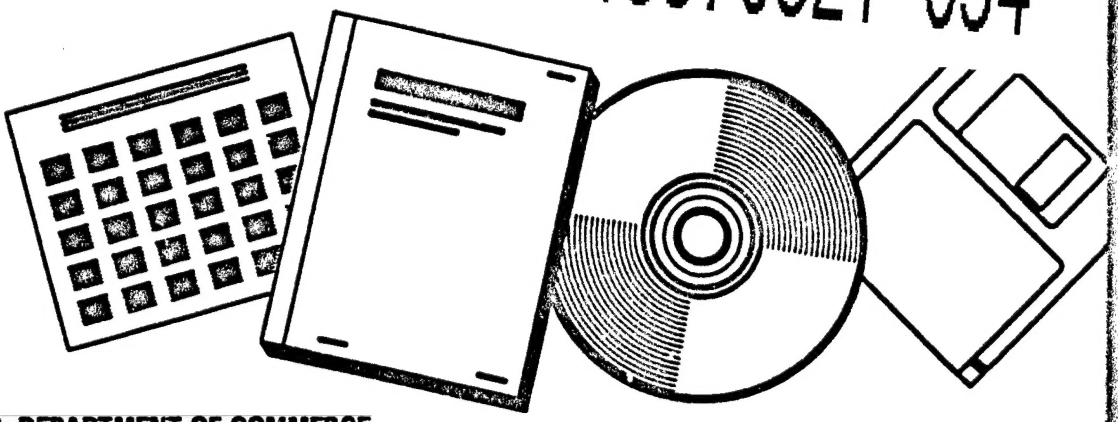
PB96-149315

WEIGHTED RANDOM MAPPINGS; PROPERTIES AND APPLICATIONS

STANFORD UNIV., CA

MAY 85



U.S. DEPARTMENT OF COMMERCE
National Technical Information Service

BIBLIOGRAPHIC INFORMATION

PB96-149315

Report Nos: STAN-CS-85-1054

Title: Weighted Random Mappings: Properties and Applications.

Date: cMay 85

Authors: A. Z. Broder.

Performing Organization: Stanford Univ., CA. Dept. of Computer Science.

Sponsoring Organization: *National Science Foundation, Washington, DC.*Office of Na Research, Arlington, VA.

Contract Nos: NSF-MCS-77-23738, NSF-DCR-83-00984, ONR-N00014-81-K-0269

Type of Report and Period Covered: Doctoral thesis.

NTIS Field/Group Codes: 62 (Computers, Control & Information Theory), 72 (Mathemati Sciences)

Price: PC A05/MF A01

Availability: Available from the National Technical Information Service, Springfiel VA. 22161

Number of Pages: 89p

Keywords: *Random functions, *Mathematical models, *Probability theory, *Computatio Runtime(Computers), Graph theory, Probability distribution functions, Permutations, Factorization, Heuristic methods, Parallel programming, Algorithms, Theses.

Abstract: A random mapping is a random graph where every vertex has outdegree one. Previous work was concerned mostly with a uniform probability distribution on these mappings. In contrast, this investigation assumed a non-uniform model, where differ mappings have different probabilities. An important application is the analysis of factorization heuristic due to Pollard and Brent. The model involved is a random mapping where every vertex has indegree either 0 or d. This distribution belongs to class called permutation invariant. A study of the general properties of permutatio invariant mappings combined with the analysis of this particular distribution made possible the computation of the expected running time of this factorization method, settling an open conjecture of Pollard.

May 1985

Report No. STAN-CS-85-1054

Weighted Random Mappings; Properties and Applications

by

Andrei Zary Broder

Department of Computer Science

Stanford University
Stanford, CA 94305

Weighted Random Mappings; Properties and Applications

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF COMPUTER SCIENCE
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

by
Andrei Zary Broder

May 1985

This research was supported in part by the National Science Foundation under grants NSF MCS 77-23738 and NSF DCR 83-00984, and by the Office of Naval Research under grant ONR N00014-81-K-0269.

Abstract

A random mapping is a random graph where every vertex has outdegree one. Previous work was concerned mostly with a uniform probability distribution on these mappings. In contrast, this investigation assumes a non-uniform model, where different mappings have different probabilities.

An important application is the analysis of a factorization heuristic due to Pollard and Brent. The model involved is a random mapping where every vertex has indegree either 0 or d. This distribution belongs to a class called *permutation invariant*. A study of the general properties of permutation invariant mappings combined with the analysis of this particular distribution made possible the computation of the expected running time of this factorization method, settling an open conjecture of Pollard.

Acknowledgments

My thanks go first and foremost to Donald Knuth. Having him as an adviser not only determined the direction of my research, but also shaped my view of Computer Science and of the meanings and goals of scholarship.

Don helped me regain my confidence in the difficult moments when months of work seemed to crumble; he was always ready to listen and advise, to suggest new avenues to explore, and to direct me to the right sources, be they ancient manuscripts in the British Museum, nineteenth century French mathematics, or yet unpublished works of his correspondents. Almost every page in this dissertation was improved by his remarks.

My gratitude also goes to the other members of my reading committee: Ernst Mayr, who besides offering substantial help over the years, also patiently read and corrected the first drafts of this dissertation, and to Andy Yao, who gave me considerable valuable advice and insights.

I would also like to acknowledge two of my former teachers that had an important influence on my academic career: Shimon Even, whose course on "Combinatorics and Graph Theory" at Technion determined me to switch major from Electrical Engineering to Computer Science, and Bob Tarjan whose classes convinced me that I wanted to specialize in the Analysis of Algorithms.

During 1983 and part of 1984 I worked at IBM Research Laboratories in San Jose as a student research associate. My interaction with Danny Dolev, Ron Fagin, Joe Halpern, Maria Klawe, Nick Pippenger, Barbara Simons, and Larry Stockmeyer, helped widen considerably my awareness of the many facets of Computer Science. Maria deserves my special thanks, not only for excellent technical advice, but also for her moral support and practical advice in reaching the right career decision.

I received valuable insights, suggestions, and references, from my discussion and correspondence with Richard Brent, Persi Diaconis, Philippe Flajolet, Dan Greene, Ehud Karnin, Andrew Odlyzko, John Pollard, Lyle Ramshaw, and Adi Shamir. The help of these friends and colleagues has resulted in a better work.

Many of the computations in this dissertation were done or checked with the help of the MAPLE symbolic manipulation system developed at the University of Waterloo.

To my parents,
Dan and Lucia Broder.

Table of contents

Page

Introduction 1

1. Permutation invariant mappings 3

1.1. Introduction and notations 3

1.2. The distribution of λ and μ 5

1.3. Higher order moments 11

1.4. The distribution of the number of cycles 13

1.5. The Foata-Fuchs encoding of mappings 15

1.6. The distribution of the repetition index 18

1.7. The distribution of ρ 19

1.8. The covariance of λ and μ 22

1.9. A simple example – permutations 23

1.10. Other types of invariance 25

2. The uniform distribution model 27

2.1. Preliminaries 27

2.1.1. The r-Stirling numbers 27

2.1.2. The Q-series 28

2.1.3. Asymptotics of certain Q-series 30

2.1.4. Abelian identities 34

2.2. The distribution of the number of cyclic elements 36

2.3. The distribution of λ and μ 38

2.4. The distribution of ρ 39

2.5. The distribution of the number of cycles 41

2.6. The transitive closure of a random mapping 43

2.6.1. Fixed starting set 43

2.6.2. Random starting points 48

2.6.3. Digression – a combinatorial identity 50

2.7. The number of ancestors of one point 51

2.8. The number of ancestors of a set of points 52

3. Pollard's factorization method 55

3.1. Pollard's factorization method 56

3.2. The constant indegree model 57

3.3. Sums of Bernoulli random variables 61

3.4. Better asymptotics 63

3.5. The case $d = 2$ 64
3.6. The parallelization of Pollard's factorization algorithm 66
3.7. Open problems 70
References 71
A. A bibliography of random mappings 75
B. A strange example 77

Table of figures

Page

1.1. Two corresponding mappings 5
1.2. A mapping on 16 elements 16
2.1. An encoding example 45

Introduction

The main objective of this dissertation is the study of finite random functions under the assumption that different functions might have different probabilities.

There exists already a relatively large literature about random functions; more than thirty journal articles in the last twenty five years have been inspired by various problems in statistics, probability theory, and biomathematics. In computer science, random mappings appear in the analysis of hashing algorithms, random number generators, cryptography problems, and integer factorization algorithms. Most of the published results concern the uniform probability distribution on the space of random functions; a uniform model is inadequate for the precise study of certain algorithms, which motivated the present work.

The first chapter introduces a certain type of probability distribution on the space of random mappings, called *permutation invariant*, that turns out to be a key concept in the analysis that follows. Using combinatorial methods it is shown that for permutation invariant distributions the random variables of interest are related in simple and unexpected ways, and it is enough to know the probability generating function for one of the variables to compute the probability generating functions for the others.

These results are applied in the second chapter to the uniform distribution model to obtain in a simple and consistent manner most of the previously known results as well as some new or sharper formulæ. Through the use of novel Abel type identities and the algebra of Q-series, all the results are expressed in terms of the Q-functions, thus explaining some of the "mysteriously" similar asymptotic behavior already noted in the literature. The last sections of chapter two do not

Chapter 1

Permutation invariant mappings

1.1. Introduction and notations

A *mapping* is a function from a set into itself. The graph of a mapping f is a directed graph where every node has outdegree one; it consists of a collection of directed trees with their roots linked into directed loops. The elements that are part of a loop are called *cyclic* or *recurrent*. (That is, an element x is recurrent iff there exists $j > 0$ such that $f^j(x) = x$.) The restriction of f to the recurrent elements is clearly a permutation of those elements.

The set of n^n functions $f : \{1,\ldots,n\} \mapsto \{1,\ldots,n\}$ is denoted $F[n]$. To each $f \in F[n]$ we associate a probability weight $w(f)$, such that $\sum_{f \in F[n]} w(f) = 1$.

A probability weight w is called *permutation invariant* if for any permutation p of $\{1,\ldots,n\}$, and for any function f, we have $w(f) = w(p \circ f)$, where the notation $p \circ f$ is defined by $(p \circ f)(x) = f\big(p(x)\big)$. An equivalent definition is that a probability weight is permutation invariant if any two functions that have as their image the same multiset, are equally likely. Indeed, consider two functions f and g such that $f\big(\{1,\ldots,n\}\big) = g\big(\{1,\ldots,n\}\big)$ as multisets. Then there exists a permutation p such that $g = p \circ f$.)

As an example of a permutation invariant distribution consider the weight

defined, for a fixed set $A \subset \{1,\ldots,n\}$ of size k, by the rule

$$w(f) = \begin{cases} 1/k^n, & \text{if } f(x) \in A \text{ for } 1 \le x \le n; \\ 0, & \text{otherwise.} \end{cases}$$

This weight is permutation invariant because for any permutation p, of $\{1,\ldots,n\}$, if $f(x) \in A$ for $1 \le x \le n$, then $f(p(x)) \in A$ for $1 \le x \le n$. More generally, if $(w_1,\ldots,w_n)$ are any weights that sum to 1, the weight function $w(f) = w_{f(1)} \cdots w_{f(n)}$ is permutation invariant.

As an example of a weight that is *not* permutation invariant consider

$$w(f) = \begin{cases} 1/n!, & \text{if } f(x) \le x \text{ for } 1 \le x \le n; \\ 0, & \text{otherwise.} \end{cases}$$

Let f be a function in $F[n]$. Define a sequence $x_{i+1} = f(x_i)$. This sequence is ultimately periodic for every x_0, and there exist numbers λ and μ, which depend on x_0, such that $x_0,\ldots,x_{\mu+\lambda-1}$ are distinct, but $x_{i+\lambda} = x_i$, for $i \ge \mu$. The number $\lambda(x_0, f)$ is called the cycle length, and the number $\mu(x_0, f)$ is called the tail length. The cycle length is always positive, but the tail length can be 0. In fact, x_0 is cyclic if and only if $\mu(x_0, f) = 0$.

Having fixed a probability weight on $F[n]$ we can define the following two random variables: $\mu(x)$ = the length of the tail starting from a certain element x and $\lambda(x)$ = the length of the period starting from x. We also define the random variables λ and μ that represent the length of the tail, and the length of the period, when f is chosen in $F[n]$ according to the probability weight w, and the starting point is chosen uniformly at random in the set $\{1,\ldots,n\}$. Hence in this case the probability space is $F[n] \times \{1,\ldots,n\}$, while in the former case it was $F[n]$ only. To avoid confusion, we shall use the notation $\mathbf{Pr}(X)$ to mean the probability of the event X when the probability space is $F[n]$ and the notation $\widehat{\mathbf{Pr}}(X)$ when the probability space is $F[n] \times \{1,\ldots,n\}$. Similar conventions apply to $\widehat{\mathbf{E}}$, $\widehat{\mathbf{var}}$, and $\widehat{\mathbf{cor}}$.

From this definition it follows that

$$\widehat{\mathbf{Pr}}(\lambda = k) = \frac{1}{n} \sum_{x \in \{1,\ldots,n\}} \mathbf{Pr}(\lambda(x) = k), \tag{1.1}$$

and similarly

$$\widehat{\mathbf{Pr}}(\mu = k) = \frac{1}{n} \sum_{x \in \{1,\ldots,n\}} \mathbf{Pr}(\mu(x) = k). \tag{1.2}$$

Yet another random variable of interest is r, the total number of cyclic elements. (The probability space for r is always $F[n]$.) The nice thing about permutation invariant weights is that there exist simple relations between the probability distributions of λ, μ, and r. We shall explore these relationships in the next sections.

1.2. The distribution of λ and μ

Lemma 1. *Given a permutation invariant probability weight w on $F[n]$, for any fixed starting point $x \in \{1,\ldots,n\}$*

$$\begin{aligned}&\Pr(\lambda(x) = i \text{ and } \mu(x) = j \mid x \text{ is not cyclic})\\ &\qquad = \Pr(\lambda(x) = j \text{ and } \mu(x) = i \mid x \text{ is not cyclic}).\end{aligned}$$

Proof: The idea of the proof is to show a 1-1 correspondence between mappings where $\lambda(x) = i$ and $\mu(x) = j$ and mappings where $\lambda(x) = j$ and $\mu(x) = i$.

Consider a mapping f such that $\lambda(x,f) = i$ and $\mu(x,f) = j$. (Because x is not cyclic, j must be strictly positive.) Consider another mapping g, identical to f everywhere except for the points x and $f^j(x)$ where $g(x) = f(f^j(x))$ and $g(f^j(x)) = f(x)$. (See Figure 1.)

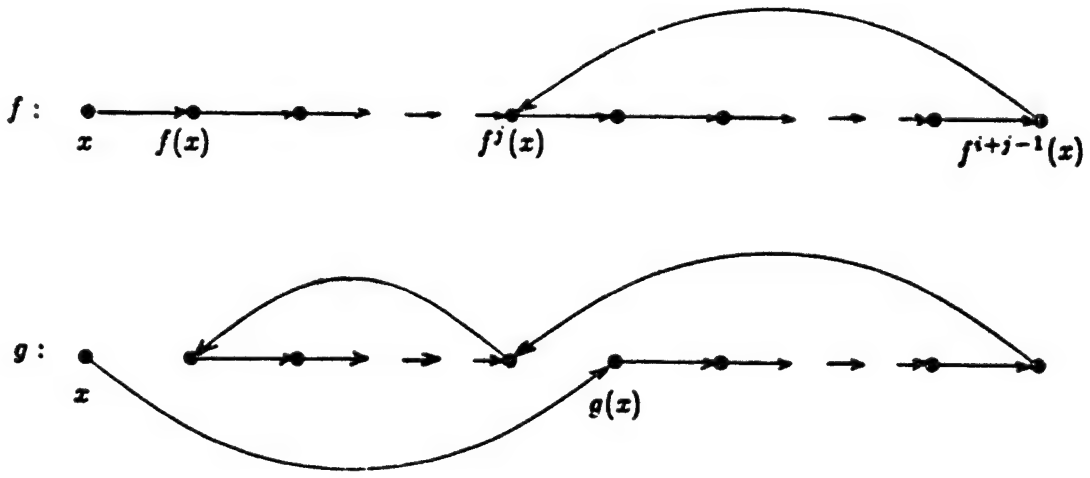


Figure 1.1. Two corresponding mappings.

It is clear that $\lambda(x,g) = j$ and $\mu(x,g) = i$. By construction $w(g) = w(f)$, because w is permutation invariant, and $g = \left(x, f^j(x)\right) \circ f$. Furthermore the correspondence $f \leftrightarrow g$ is one to one, and the desired result is obtained by summing the probabilities of all f's with $\lambda(x,f) = i$ and $\mu(x,f) = j$. ∎

Corollary 2. *Given a permutation invariant probability weight w on $F[n]$, for any starting point $x \in \{1,\ldots,n\}$*

$$\mathbf{Pr}\big(\mu(x) = i \mid x \text{ is not cyclic}\big) = \mathbf{Pr}\big(\lambda(x) = i \mid x \text{ is not cyclic}\big).$$

∎

Corollary 3. *Given a permutation invariant probability weight w on $F[n]$, for any starting point $x \in \{1,\ldots,n\}$, and for any $i > 0$*

$$\mathbf{Pr}\big(\mu(x) = i\big) = \mathbf{Pr}\big(\lambda(x) = i\big) - \mathbf{Pr}\big(\lambda(x) = i \text{ and } x \text{ is cyclic}\big).$$

Proof: Because i is positive

$$\mathbf{Pr}\big(\mu(x) = i\big) = \mathbf{Pr}\big(\mu(x) = i \text{ and } x \text{ is not cyclic}\big);$$

Using the Bayes rule and Corollary 2, we obtain

$$\begin{aligned}\mathbf{Pr}\big(\mu(x) = i\big) &= \mathbf{Pr}\big(\mu(x) = i \text{ and } x \text{ is not cyclic}\big)\\ &= \mathbf{Pr}\big(x \text{ is not cyclic}\big)\,\mathbf{Pr}\big(\mu(x) = i \mid x \text{ is not cyclic}\big)\\ &= \mathbf{Pr}\big(x \text{ is not cyclic}\big)\,\mathbf{Pr}\big(\lambda(x) = i \mid x \text{ is not cyclic}\big)\\ &= \mathbf{Pr}\big(\lambda(x) = i \text{ and } x \text{ is not cyclic}\big).\end{aligned}$$

∎

For several of the following proofs it is convenient to define an equivalence relation on $F[n]$ as follows. Two mappings $f, g \in F[n]$ will be called *similar* if they have the same set of cyclic elements, $A \subset \{1,\ldots,n\}$, and $f(x) = g(x)$ for all $x \notin A$. That means that the set of mappings similar to f is obtained by composing an arbitrary permutation of the cyclic elements of f with f itself. Clearly similarity is an equivalence relation, and the set of all the equivalence classes under it will be denoted $E[n]$.

Observe that

- If f belongs to some class $E \in E[n]$ and f has k cyclic elements then $|E| = k!$, and hence the only possible values for the cardinality of an arbitrary equivalence class are $1!, 2!, \ldots$. (The number of equivalence classes of size $k!$ will be discussed in Section 6.)
- If f and g belong to the same equivalence class and the weight w is permutation invariant, then $w(f) = w(g)$.

Lemma 4. *Given a permutation invariant probability weight w on $F[n]$ and an integer $1 \le i \le n$, if x is chosen uniformly at random, then*

$$\widehat{\Pr}(\lambda(x) = i \text{ and } x \text{ is cyclic}) = \frac{1}{n} \sum_{k \ge i} \Pr(r = k).$$

Proof: Fix one similarity equivalence class; all the mappings in it have the same cyclic elements. Let k be the number of these cyclic elements. The probability that x is among them is k/n, and the probability that the cycle containing x has length $i \le k$ is exactly $1/k$ because all the permutations of the cyclic elements are equally likely (see [Knuth73a, ex. 1.3.3–17]). Summing on all possible sets of k cyclic elements and on all $k \ge i$ completes the proof; more formally we have

$$\begin{aligned}
&\widehat{\Pr}(\lambda(x) = i \text{ and } x \text{ is cyclic}) \\
&\qquad = \frac{1}{n} \sum_{f \in F[n]} \sum_{1 \le x \le n} w(\{f \mid \lambda(x, f) = i \text{ and } x \text{ cyclic in } f\}) \\
&\qquad = \frac{1}{n} \sum_{i \le k \le n} \sum_{|E| = k} \sum_{f \in E} \sum_{1 \le x \le n} w(\{f \mid \lambda(x, f) = i \text{ and } x \text{ cyclic in } f\}) \\
&\qquad = \sum_{i \le k \le n} \sum_{|E| = k} \frac{k}{n} \frac{1}{k} w(\{f \mid f \in E\}) = \frac{1}{n} \sum_{i \le k \le n} \Pr(r = k).
\end{aligned}$$

A somewhat surprising property of the permutation invariant mappings follows directly from the above lemma:

Corollary 5. *For any permutation invariant probability weight w on $F[n]$ the expected number of fixed points is 1.*

Proof: From Lemma 4, if x is chosen uniformly at random then

$$\widehat{\Pr}(\lambda(x) = 1 \text{ and } x \text{ is cyclic}) = \frac{1}{n} \sum_{k \ge 1} \Pr(r = k) = \frac{1}{n},$$

since every function has at least one cyclic element. On the other hand, if x is chosen uniformly at random then

$$\begin{aligned}
\widehat{\Pr}(\lambda(x) = 1 \text{ and } x \text{ is cyclic}) &= \frac{1}{n} \sum_{x \in \{1, \ldots, n\}} \Pr(\lambda(x) = 1 \text{ and } x \text{ is cyclic}) \\
&= \frac{1}{n} \mathrm{E}(\text{number of fixed points}).
\end{aligned}$$

∎

Lemma 6. *Given a permutation invariant probability weight w on $F[n]$, for any fixed point x and any fixed integer $1 \le i \le n$,*

$$\Pr(\lambda(x) = i) = \sum_{k \ge i} \frac{\Pr(r = k)}{k}.$$

Proof: Consider a certain mapping f such that y is the first cyclic point reached from x. Within the equivalence class of f, all permutations induced by the cyclic points are equally likely, and therefore the length of the cycle containing y is uniformly distributed between 1 and k. The proof is completed by summing over all the equivalence classes, weighted by their probability, as in the proof of Lemma 4. ∎

Corollary 7. *Given a permutation invariant probability weight w on $F[n]$, for all $i > 0$,*

$$\widehat{\Pr}(\mu = i) = \widehat{\Pr}(\lambda = i) - \sum_{k \ge i} \frac{\Pr(r = k)}{n} = \sum_{k \ge i} \Pr(r = k) \left(\frac{1}{k} - \frac{1}{n} \right).$$

Proof: By Corollary 3 and Lemma 6, we have

$$\Pr\big(\mu(x) = i\big) = \Pr\big(\lambda(x) = i\big) - \Pr\big(\lambda(x) = i \text{ and } x \text{ is cyclic}\big),$$

for any fixed x, and therefore also for x chosen uniformly at random, in which case we can also apply Lemma 4. ∎

We are now ready to derive the relations between the distribution of the number of cyclic points, r, and the distribution of λ and μ.

Theorem 8. *Let the probability generating functions for λ, μ, and r be $L(z)$, $M(z)$, and $C(z)$, respectively. Given any permutation invariant probability weight w on $F[n]$, these functions satisfy*

$$L(z) = \frac{z}{z-1} \int_1^z \frac{C(t)}{t}\, dt,$$

and

$$M(z) = \frac{C'(1)}{n} + L(z) - \frac{z\big(C(z) - 1\big)}{n(z-1)}.$$

Proof: By definition and Lemma 6 applied to a starting point chosen uniformly at random,

$$L(z) = \sum_i \widehat{\Pr}(\lambda = i) z^i = \sum_{k \geq 1} \frac{\Pr(\tau = k)}{k} \sum_{1 \leq i \leq k} z^i$$
$$= z \sum_{k \geq 1} \frac{\Pr(\tau = k)}{k} \frac{z^k - 1}{z - 1}.$$

On the other hand

$$\int_1^z \frac{C(t)}{t}\, dt = \sum_{k \geq 1} \frac{\Pr(\tau = k) t^k}{k} \Big|_1^z = \sum_{k \geq 1} \frac{\Pr(\tau = k)}{k} (z^k - 1),$$

which proves the first part of the theorem.

For the second part we use Corollary 7, from which we obtain

$$\sum_{i \geq 1} \widehat{\Pr}(\mu = i) z^i = L(z) - \sum_{i \geq 1} z^i \sum_{k \geq i} \frac{\Pr(\tau = k)}{n} = L(z) - \sum_{k \geq 1} \frac{\Pr(\tau = k)}{n} \sum_{1 \leq i \leq k} z^i$$
$$= L(z) - \frac{z}{n(z-1)} \sum_{k \geq 1} \Pr(\tau = k)(z^k - 1) = L(z) - \frac{z(C(z) - 1)}{n(z-1)}.$$

The probability that $\mu = 0$ is the same as the probability of choosing a cyclic element,

$$\widehat{\Pr}(\mu = 0) = \sum_k \Pr(\tau = k) \frac{k}{n} = \frac{C'(1)}{n}.$$

Combining the last two equations we get

$$M(z) = \frac{C'(1)}{n} + L(z) - \frac{z(C(z) - 1)}{n(z-1)}.$$

∎

In a similar manner, it is possible to obtain expressions for $C(z)$ and $L(z)$ in terms of $M(z)$, or for $M(z)$ and $C(z)$ in terms of $L(z)$.

As a quick check, note that by L'Hospital's rule

$$\lim_{z \to 1} L(z) = \lim_{z \to 1} \frac{C(z)}{z} = C(1) = 1,$$

and also

$$\lim_{z \to 1} M(z) = \frac{1}{n} C'(1) + \lim_{z \to 1} L(z) - \frac{1}{n} \lim_{z \to 1} C'(z) = 1.$$

Using Theorem 8 we can easily compute the means of λ and μ in terms of the moments of τ, but it is more fun to prove their relations directly.

Theorem 9. *Given a permutation invariant probability weight w on $F[n]$*

$$\hat{\mathrm{E}}(\lambda) = \frac{\mathrm{E}(\tau)+1}{2}.$$

Proof: Choose a fixed starting point x. Let y be the first cyclic point reached from x. By the argument used in the proof of Lemma 6, the expected length of the cycle containing y averaged over all the mappings with k cyclic elements is $(k+1)/2$. Hence

$$\mathrm{E}(\lambda(x)) = \frac{\mathrm{E}(\tau)+1}{2}.$$

∎

Theorem 10. *Given a permutation invariant probability weight w on $F[n]$*

$$\hat{\mathrm{E}}(\mu) = \hat{\mathrm{E}}(\lambda) - \frac{1}{2n}\big(\mathrm{E}(\tau)+\mathrm{E}(\tau^2)\big).$$

Proof: If the starting point is not cyclic the expected value of λ and μ are the same (Corollary 2); therefore to obtain the mean value of μ we must subtract from the mean value of λ the contribution of the cyclic elements, which have $\mu = 0$. Assuming k cyclic elements, their total contribution is $k(k+1)/2$, so that the contribution per element is $k(k+1)/(2n)$ and the claim follows. ∎

To obtain the same relations from Theorem 8, we compute

$$\begin{aligned}
L'(1) &= \lim_{z\to 1}\left(\frac{z}{z-1}\frac{C(z)}{z} - \frac{1}{(z-1)^2}\int_1^z \frac{C(t)}{t}\,dt\right)\\
&= \lim_{z\to 1}\frac{(z-1)C(z) - \int_1^z C(t)/t\,dt}{(z-1)^2}\\
&= \lim_{z\to 1}\frac{C(z)+(z-1)C'(z) - C(z)/z}{2(z-1)}\\
&= \lim_{z\to 1}\left(\frac{1}{2}C'(z) + \frac{1}{2z}C(z)\right) = \frac{C'(1)+C(1)}{2},
\end{aligned}$$

and similarly

$$\begin{aligned}
M'(1) &= L'(1) - \frac{1}{n}\lim_{z\to 1}\left(\frac{zC'(z)}{z-1} - \frac{C(z)-1}{(z-1)^2}\right)\\
&= L'(1) - \frac{1}{n}\lim_{z\to 1}\frac{z(z-1)C'(z) - C(z) + 1}{(z-1)^2}\\
&= L'(1) - \frac{1}{n}\lim_{z\to 1}\frac{(z-1)C'(z) + zC'(z) + z(z-1)C''(z) - C'(z)}{2(z-1)}\\
&= L'(1) - \frac{1}{2n}\big(C''(1) + 2C'(1)\big).
\end{aligned}$$

We can obtain higher order moments in the same manner but it is more convenient to use the approach described in the next section.

1.3. Higher order moments

Given a probability distribution $G(z) = \sum_k g_k z^k$, its lth *factorial moment*, $G^{(l)}$, is defined by

$$G^{(l)} = \frac{d^l}{dz^l} G(z)\Big|_{z=1} = \sum_k g_k k^{\underline{l}}.$$

Our goal in this section is to express the higher order factorial moments of λ and μ using the higher order factorial moments of τ, namely $C^{(l)} = \sum_k c_k k^{\underline{l}}$, where $c_k = \mathbf{Pr}(\tau = k)$.

The following discussion is simplified if we use the notation

$$G_l = \frac{G^{(l)}}{l!} = \frac{1}{l!}\frac{d^l}{dz^l} G(z)\Big|_{z=1} = \sum_k g_k \binom{k}{l}.$$

By Taylor's theorem, if all the moments of G exist, then

$$\begin{aligned} G(w+1) &= G(1) + \frac{wG'(1)}{1!} + \frac{w^2 G''(1)}{2!} + \cdots \\ &= G_0 + wG_1 + w^2 G_2 + \cdots. \end{aligned} \tag{1.3}$$

From Theorem 8 we have

$$\begin{aligned} L(w+1) &= \frac{w+1}{w}\int_1^{w+1} \frac{C(t)}{t}\,dt = \frac{w+1}{w}\int_0^{w} \frac{C(t+1)}{t+1}\,dt \\ &= \frac{w+1}{w}\int_0^{w} (C_0 + C_1 t + C_2 t^2 + \cdots)(1 - t + t^2 + \cdots)\,dt \\ &= \frac{w+1}{w}\int_0^{w} (C_0 + (C_1 - C_0)t + (C_2 - C_1 + C_0)t^2 + \cdots)\,dt \\ &= (w+1)\left(C_0 + \frac{1}{2}(C_1 - C_0)w + \frac{1}{3}(C_2 - C_1 + C_0)w^2 + \cdots\right). \end{aligned} \tag{1.4}$$

Hence for $l > 0$

$$\begin{aligned} L_l &= \frac{1}{l+1}\sum_{0\le i\le l} (-1)^{l-i} C_i + \frac{1}{l}\sum_{0\le i\le l-1} (-1)^{l-1-i} C_i \\ &= \frac{C_l}{l+1} - \frac{1}{l(l+1)}\sum_{0\le i\le l-1} (-1)^{l-i} C_i. \end{aligned} \tag{1.5}$$

In particular

$$L'(1) = \frac{1}{2}\big(C'(1) + 1\big), \tag{1.6}$$

which we already know (Theorem 9), and

$$L''(1) = \frac{1}{3}\big(C''(1) + C'(1) - 1\big). \tag{1.7}$$

Some more formula crunching, better left to a computer, leads to

$$\begin{aligned} \widehat{\mathrm{var}}(\lambda) &= L''(1) + L'(1) - \big(L'(1)\big)^2 \\ &= \frac{1}{3}\,\mathrm{var}(\tau) + \frac{1}{12}\,\mathrm{E}(\tau)^2 - \frac{1}{12}. \end{aligned} \tag{1.8}$$

For the moments of μ, first concentrate on

$$A(z) = \frac{z\big(C(z) - 1\big)}{n(z-1)}. \tag{1.9}$$

We have

$$A(w+1) = \frac{(w+1)\big(C(w+1) - 1\big)}{nw} = \frac{w+1}{n}\big(C_1 + C_2 w + C_3 w^2 + \cdots\big), \tag{1.10}$$

hence

$$A_l = \frac{C_l + C_{l+1}}{n}. \tag{1.11}$$

On the other hand

$$M(z) = \frac{C'(1)}{n} + L(z) - A(z),$$

so that finally

$$M_l = L_l - \frac{C_l + C_{l+1}}{n}, \qquad l > 0. \tag{1.12}$$

In particular

$$M'(1) = L'(1) - \frac{C''(1) + 2C'(1)}{2n}, \tag{1.13}$$

and

$$M''(1) = L''(1) - \frac{C'''(1) + 3C''(1)}{3n}. \tag{1.14}$$

1.4. The distribution of the number of cycles

Another interesting characteristic of a mapping is the number of components, or equivalently, the number of cycles. This random variable (over $F[n]$) is denoted β, and its probability generating function is denoted $B(z)$. We shall see below that the distribution of β is closely related to the distribution of τ.

Theorem 11. *Given a permutation invariant probability weight w on $F[n]$, the probability distribution of the number of components satisfies*

$$\mathbf{Pr}(\beta = j) = \sum_k \begin{bmatrix} k \\ j \end{bmatrix} \frac{1}{k!} \mathbf{Pr}(\tau = k).$$

Proof: Fix k cyclic elements. All their permutations are equally likely, hence the probability of their forming j cycles is exactly $\begin{bmatrix} k \\ j \end{bmatrix}/k!$, where $\begin{bmatrix} k \\ j \end{bmatrix}$ is a signless Stirling number of the first kind[1] (See, for example, [Knuth73a, §1.2.10].) ∎

Corollary 12. *Given a permutation invariant probability weight w on $F[n]$, the probability of a function being connected is*

$$\mathbf{Pr}(\beta = 1) = \sum_k \frac{\mathbf{Pr}(\tau = k)}{k} = \int_0^1 \frac{C(z)}{z} dz.$$

∎

From Theorem 11, the probability generating function of β satisfies

$$B(z) = \sum_k c_k \sum_j \begin{bmatrix} k \\ j \end{bmatrix} \frac{z^j}{k!} = \sum_j z^j \sum_k \begin{bmatrix} k \\ j \end{bmatrix} \frac{c_k}{k!}. \tag{1.15}$$

On the other hand the exponential generating function for Stirling numbers of the first kind is

$$\sum_k \begin{bmatrix} k \\ j \end{bmatrix} \frac{t^k}{k!} = \frac{1}{j!} \left(\ln \frac{1}{1-t} \right)^j,$$

and therefore, using the Hadamard product,

$$\sum_k \begin{bmatrix} k \\ j \end{bmatrix} \frac{c_k}{k!} = \frac{1}{j!} \frac{1}{2\pi i} \oint \frac{1}{t} C(t) \left(\ln \frac{t}{t-1} \right)^j dt, \tag{1.16}$$

[1] With this notation: $k! \binom{z}{k} = \sum_j (-1)^{k-j} \begin{bmatrix} k \\ j \end{bmatrix} z^j$.

where the integral is on a circle around the origin.

Combining equations (15) and (16), we obtain

$$B(z) = \sum_j \frac{z^j}{j!} \frac{1}{2\pi i} \oint \frac{1}{t} C(t) \left(\ln \frac{t}{t-1} \right)^j dt$$
$$= \frac{1}{2\pi i} \oint \frac{1}{t} C(t) \left(\frac{t}{t-1} \right)^z dt \qquad (1.17)$$

To compute the moments of β we differentiate equation (17) with respect to z, at $z = 1$. We obtain

$$l!\, B_l = \frac{1}{2\pi i} \oint \frac{1}{t} C(t) \left(\ln \frac{t}{t-1} \right)^l \frac{t}{t-1} dt;$$

Now we use the expansion

$$\frac{1}{l!} \left(\frac{1}{1-z} \right)^r \left(\ln \frac{1}{1-z} \right)^l = \sum_k \begin{bmatrix} k+r \\ l+r \end{bmatrix}_r \frac{z^k}{k!},$$

where $\begin{bmatrix} k+r \\ l+r \end{bmatrix}_r$ is an r-Stirling numbers of the first kind.[2] (see [Broder84a]). Setting $z = 1/t$ and $r = 1$ we get

$$\frac{t}{t-1} \left(\ln \frac{t}{t-1} \right)^l = \sum_k \frac{l!}{k!} \begin{bmatrix} k+1 \\ l+1 \end{bmatrix}_1 t^{-k}.$$

For $l > 0$, $\begin{bmatrix} k+1 \\ l+1 \end{bmatrix}_1 = \begin{bmatrix} k+1 \\ l+1 \end{bmatrix}$, and we obtain

$$B_l = \sum_k \frac{c_k}{k!} \begin{bmatrix} k+1 \\ l+1 \end{bmatrix}, \qquad l > 0. \qquad (1.18)$$

For the first moments it might be preferable to use the identities ([Zave76])

$$\frac{1}{1-z} \ln \frac{1}{1-z} = \sum_k H_k z^k,$$

[2] The r-Stirling numbers of the first kind, $\begin{bmatrix} k \\ l \end{bmatrix}_r$, count the number of permutations of k elements, with l cycles, such that the elements $1, 2, \ldots, r$ are in distinct cycles. The r-Stirling numbers of the second kind are discussed in section 2.1.1.

and

$$\frac{1}{1-z}\left(\ln\frac{1}{1-z}\right)^2 = \sum_k (H_k^2 - H_k^{(2)})z^k,$$

where $H_k^{(l)}$ are (generalized) Harmonic numbers, $H_k = \sum_{1\le j\le k} j^{-1}$, and $H_k^{(2)} = \sum_{1\le j\le k} j^{-2}$.

From these identities

$$\mathbf{E}(\beta) = \sum_k c_k H_k, \tag{1.19}$$

and

$$\mathbf{E}\big(\beta(\beta-1)\big) = \sum_k c_k (H_k^2 - H_k^{(2)}). \tag{1.20}$$

We have seen so far that the distributions of several important random variables associated with a permutation invariant probability distribution on $F[n]$, are determined by the distribution of r, the number of cyclic elements. The next two sections show how this distribution can be replaced by the distribution of a simpler entity.

1.5. The Foata-Fuchs encoding of mappings

Let $\Sigma = \{\bar{1}, \bar{2}, \ldots, \bar{n}\}$ be an alphabet. Using the terminology of Comtet [Comtet74], to each word ω over Σ we associate its *Abelian image*, $\mathcal{T}(\omega)$ obtained by replacing each occurrence of the letter $\bar{\imath}$ in ω by the commutative variable x_i. For instance $\mathcal{T}(\bar{1}\,\bar{2}\,\bar{3}\,\bar{2}) = x_1x_2^2x_3$. The *enumerator* of a multiset of words $\Omega \subset \Sigma^*$ is defined to be the polynomial

$$\mathcal{P}(\Omega) = \sum_{\omega\in\Omega} \mathcal{T}(\omega).$$

Clearly the coefficient of $x_1^{i_1}x_2^{i_2}\ldots x_n^{i_n}$ in $\mathcal{P}(\Omega)$ represents the number of words in Ω that contain exactly i_1 occurrences of $\bar{1}$, i_2 occurrences of $\bar{2}$, and so on.

Sometimes only some letters are of interest; in this case the variables corresponding to other letters are assigned the value 1, to obtain the enumerator by the number of occurrences of the distinguished letters. This is the same as the enumerator of the multiset obtained from Ω by deleting from each word the undistinguished letters. For instance the enumerator of Σ^n is $(x_1 + x_2 + \cdots + x_n)^n$; the enumerator of Σ^n by the number of occurrences of $\bar{1}$ and $\bar{3}$ is $(x_1 + x_3 + n - 2)^n$.

An *encoding* is a 1-1 correspondence between the n^n distinct mappings from $\{1,2,\ldots,n\}$ into $\{1,2,\ldots,n\}$ and the n^n distinct words of length n over the alphabet $\Sigma = \{\bar{1}, \bar{2}, \ldots, \bar{n}\}$. Let f be a mapping. The trivial encoding of f is given by

$$f \leftrightarrow \overline{f(1)}\,\overline{f(2)}\ldots\overline{f(n)};$$

that is, the word associated to f is obtained by concatenating the letters corresponding to the values taken by f in $1, 2, \ldots, n$. The trivial encoding of the mapping in Figure 2 is

$$\overline{4}\ \overline{3}\ \overline{15}\ \overline{15}\ \overline{6}\ \overline{3}\ \overline{16}\ \overline{13}\ \overline{9}\ \overline{4}\ \overline{10}\ \overline{7}\ \overline{16}\ \overline{9}\ \overline{5}\ \overline{7}.$$

The trivial encoding of a mapping f is denoted $C_0(f)$.

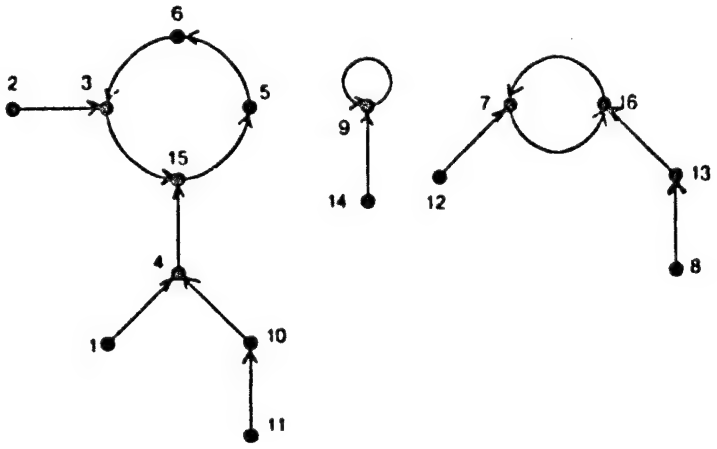


Figure 1.2. A mapping on 16 elements.

As described below, the Foata-Fuchs encoding (FF-encoding) [FF70] of a mapping f is the concatenation of $l+1$ words, $\omega_0, \omega_1, \ldots, \omega_l$ where l is the number of leaves (nodes with indegree 0) in the graph of f. The word ω_0 describes the permutation induced by the cyclic elements and is generated by the following algorithm. (After each step, the result of this algorithm, applied to the mapping in Figure 2, is shown in square brackets.)

Algorithm A.

1. Write the permutation as a product of cycles.
   $[(3,15,5,6)(9)(7,16)]$
2. Reverse each cycle.
   $[(6,5,15,3)(9)(16,7)]$
3. Rotate each cycle such that the maximum element in each cycle is in the first position. The element in the first position of a cycle is called the cycle leader.
   $[(15,3,6,5)(9)(16,7)]$
4. Put the cycles in increasing order of their cycle leader.

$$[(9)(15,3,6,5)(16,7)]$$

5. Remove parentheses and replace each number by its corresponding letter.
$$[\overline{9}\ \overline{15}\ \overline{3}\ \overline{6}\ \overline{5}\ \overline{16}\ \overline{7}]$$

□

The result is actually another permutation of the letters of $C_0(f)$. It is clear that the transformation is 1-1 because the algorithm can be reversed. The end of a cycle is "signalled" by a new left-to-right maximum. This transformation is implied in [Riordan58, Chap. 8] and formalized and generalized in [Foata65]. It was later referred to as "the first fundamental transformation for permutations" [Foata83].

Now let $a_1, a_2, \ldots, a_l$, $a_1 < a_2 < \cdots < a_l$ be the leaves in the graph of f. The word ω_i consists of the labels of the nodes on the path from a_i to the first node already appearing in $\omega_0, \ldots, \omega_{i-1}$, including this node and excluding a_i, in reverse order. For the mapping in Figure 2 we have $l = 6$, $a_1 = 1$, $a_2 = 2$, $a_3 = 8$, $a_4 = 11$, $a_5 = 12$, $a_6 = 14$, and $\omega_1 = \overline{15}\ \overline{4}$, $\omega_2 = \overline{3}$, $\omega_3 = \overline{16}\ \overline{13}$, $\omega_4 = \overline{4}\ \overline{10}$, $\omega_5 = \overline{7}$, $\omega_6 = \overline{9}$. Therefore the complete encoding of the function in Figure 2 is

$$\overline{9}\ \overline{15}\ \overline{3}\ \overline{6}\ \overline{5}\ \overline{16}\ \overline{7}\ \overline{15}\ \overline{4}\ \overline{3}\ \overline{16}\ \overline{13}\ \overline{4}\ \overline{10}\ \overline{7}\ \overline{9}\,.$$

It is clear from its definition that the FF-encoding is a permutation of the trivial encoding. It is a 1-1 correspondence because it can be reversed. Given a word ω of length n, the letters corresponding to $a_1, a_2, \ldots, a_l$ are exactly those letters among $\overline{1}, \ldots, \overline{n}$ that do not appear in ω, sorted in increasing order. The subword ω_0 ends before the first repeated letter, or $\omega_0 = \omega$ if no letter is repeated; the beginning of each subword describing a path is "signalled" by a repeated letter; and so on. Exactly l letters are repeated since l letters are left out.

Given a mapping f its FF-encoding will be denoted $C_1(f)$. A set of mappings F encodes into a set of words, denoted $C_1(F)$.

Examples

1. Let F be the family of functions $f : \{1, \ldots, n\} \to \{1, \ldots, n\}$ such that the graph of f is connected and n is cyclic. Then
$$P(C_1(F)) = x_n(x_1 + x_2 + \cdots + x_n)^{n-1},$$
because any $f \in F$ has exactly one cycle, and n must be in it. Therefore $C_1(f)$ starts with $\overline{n}$ followed by an arbitrary word of length $n-1$.

2. Let F be the family of functions $f : \{1,\dots,n\} \to \{1,\dots,n\}$ such that n is the unique cyclic element. (The graph of every f is a tree rooted at n.) Then

$$P(C_1(F)) = x_n x_n (x_1 + x_2 + \cdots + x_n)^{n-2}.$$

Making

$$x_1 = x_2 = \cdots = x_n = 1,$$

we find that the number of undirected trees on n labelled vertices is n^{n-2}. (Cayley's theorem)

1.6. The distribution of the repetition index

The repetition index ν of a mapping f is the maximum number i such that the values $f(1), f(2), \dots, f(i)$ are all distinct. For example, the mapping in Figure 2 has $\nu = 3$. Our interest in the repetition index is motivated by the next theorem.

Theorem 13. *For any permutation invariant probability weight w on $F[n]$, the probability distributions of the total number of cyclic elements and of the repetition index are equal, that is for any k*

$$\mathbf{Pr}(r = k) = \mathbf{Pr}(\nu = k).$$

Proof: Let f be a mapping with $r(f) = k$. If we look at the FF-encoding of f as the trivial encoding of another mapping g (i.e. $C_0(g) = C_1(f)$), then g has the property that $w(g) = w(f)$ because w is permutation invariant. Furthermore $\nu(g) = k$ by construction. Summing over all f with $r(f) = k$ completes the proof. ∎

Theorem 13 is the basic tool for the computation of all the probability distributions implied by a certain permutation invariant weight because in most cases it is easy to determine the distribution of ν, using simple string counting arguments.

Using the FF-encoding we can count the number of similarity equivalence classes introduced in Section 2.

Theorem 14. *The number of similarity equivalence classes of size $k!$ is*

$$\binom{n}{k} k n^{n-k-1},$$

and the total number of equivalence classes is

$$|E[n]| = (n+1)^{n-1}.$$

Proof: Consider two mappings, f and g, belonging to the same equivalence class, E. Assume that $|E| = k!$. (Hence f and g have k cyclic elements.) Under this assumption, if $C_1(f) = a_1 a_2 \dots a_n$ and $C_1(g) = b_1 b_2 \dots b_n$, we must have $a_j = b_j$ for all j greater than k; $a_1 a_2 \dots a_k$ must be all distinct and must be a permutation of $b_1 b_2 \dots b_k$; and a_{k+1} must be equal to one of the letters $a_1, a_2, \dots, a_k$. It follows that any equivalence class with $k!$ elements can be represented by a set of k distinct letters and a string of $n-k$ arbitrary letters, such that the first letter of the string is in the set. Therefore the number of equivalence classes of size $k!$ is $\binom{n}{k} k n^{n-k-1}$. The total number of equivalence classes is

$$\sum_{k \geq 1} \binom{n}{k} k n^{n-k-1},$$

which equals $(n+1)^{n-1}$ because of the identity

$$\sum_{k \geq 1} \binom{n}{k} k x^{k-1} y^{n-k} = n(x+y)^{n-1},$$

which is obtained by taking the derivative of $(x+y)^n$ with respect to x. ∎

1.7. The distribution of ρ

In certain cases we are interested in the distribution of $\rho(x, f) = \lambda(x, f) + \mu(x, f)$, that is, the sum of the length of the tail and the length of the period starting from x in the mapping f. Another interpretation for $\rho(x, f)$ is to see it as the number of elements reachable from x, or the number of "descendants" of x, in the graph of f. By analogy with λ and μ, the random variable $\rho(x)$ represents the value of $\rho(x, f)$ when x is fixed and f is chosen in $F[n]$ according to the probability weight w, and the random variable ρ represents the value of $\rho(x)$ when x is chosen uniformly at random in $\{1, \dots, n\}$.

We know the expected values of λ and of μ, hence the expected value of ρ is not hard to find. But its higher order moments present more difficulties. Fortunately, we can relate the distribution of ρ directly to the distribution of the total number of cyclic elements, r, with the help of yet another encoding of mappings.

Let $f \in F[n]$ be a mapping, and let x be some fixed element in $\{1,\ldots,n\}$. The encoding $C_2(x,f)$ is a string of length $n+1$ over the alphabet $\{\overline{1},\overline{2},\ldots,\overline{n}\}$ of the form

$$\overline{x}\ \overline{f(x)}\ \overline{f^2(x)}\ \ldots\ \overline{f^i(x)}\ \overline{f(a_1)}\ \ldots\ \overline{f(a_{n-i})}\ ,$$

where $f^i(x)$ is the first repeated element in the sequence $x, f(x), f^2(x), \ldots$, possibly x itself, and $a_1,\ldots,a_{n-i}$ are the elements in $\{1,\ldots,n\}$ that do not belong to this sequence, written in increasing order. The index i is in fact $\rho(x,f)$. For example, for the function in Figure 2 we have

$$C_2(1,f) = \overline{1}\ \overline{4}\ \overline{15}\ \overline{5}\ \overline{6}\ \overline{3}\ \overline{15}\ \overline{3}\ \overline{16}\ \overline{13}\ \overline{9}\ \overline{4}\ \overline{10}\ \overline{7}\ \overline{16}\ \overline{9}\ \overline{7},$$
$$C_2(5,f) = \overline{5}\ \overline{6}\ \overline{3}\ \overline{15}\ \overline{5}\ \overline{4}\ \overline{3}\ \overline{15}\ \overline{16}\ \overline{13}\ \overline{9}\ \overline{4}\ \overline{10}\ \overline{7}\ \overline{16}\ \overline{9}\ \overline{7},$$
$$C_2(9,f) = \overline{9}\ \overline{9}\ \overline{4}\ \overline{3}\ \overline{15}\ \overline{15}\ \overline{6}\ \overline{3}\ \overline{16}\ \overline{13}\ \overline{4}\ \overline{10}\ \overline{7}\ \overline{16}\ \overline{9}\ \overline{5}\ \overline{7}.$$

Lemma 15. *Given a permutation invariant probability weight w on $F[n]$, if*

$$C_2(x,f) = \overline{x}\ C_0(g)$$

then

$$\mathbf{Pr}(f) = \mathbf{Pr}(g).$$

Proof: Clearly, $C_2(x,f)$ consists of $\overline{x}$ followed by a permutation of the trivial encoding of f. The premise of the Lemma means that the trivial encoding of g is a permutation of the trivial encoding of f; hence f and g have the same probability because w is permutation invariant. ∎

Theorem 16. *Given a permutation invariant probability weight w on $F[n]$,*

$$\widehat{\mathbf{Pr}}(\rho = k) = \left(1 - \frac{k-1}{n}\right)\mathbf{Pr}(\tau = k-1) + \frac{1}{n}\mathbf{Pr}(\tau \geq k).$$

Proof: We have

$$\widehat{\mathbf{Pr}}(\rho = k) = \widehat{\mathbf{Pr}}\big(\rho(x) = k \text{ and } x \text{ is not cyclic}\big) + \widehat{\mathbf{Pr}}\big(\rho(x) = k \text{ and } x \text{ is cyclic}\big).$$

The latter term is just $\widehat{\mathbf{Pr}}\big(\lambda(x) = k \text{ and } x \text{ is cyclic}\big)$, which equals $\mathbf{Pr}(\tau \geq k)/n$ by Lemma 4. The former term is $1/n$ times the total weight of all mappings whose C_2-encoding has the form $xC_0(g)$, where g is a mapping such that $\nu(g) = k-1$ and x is in $\{1,\ldots,n\}$, but it is *not* one of the distinct elements $g(1),\ldots,g(k-1)$. Hence we have

$$\widehat{\mathbf{Pr}}(\rho = k) = \frac{n-k+1}{n}\mathbf{Pr}(\nu = k-1) + \frac{1}{n}\mathbf{Pr}(\tau \geq k),$$

and the theorem follows because the distributions of ν and τ are identical (Theorem 13). ∎

As a quick check we can compute

$$\sum_k \widehat{\Pr}(\rho = k) = \sum_k \left(1 - \frac{k-1}{n}\right) \Pr(\tau = k-1) + \frac{1}{n} \sum_k \Pr(\tau \geq k)$$
$$= \sum_k \left(1 - \frac{k}{n}\right) \Pr(\tau = k) + \frac{1}{n} \mathrm{E}(\tau) = 1$$

and

$$\widehat{\mathrm{E}}(\rho) = \sum_k k \left(1 - \frac{k-1}{n}\right) \Pr(\tau = k-1) + \frac{1}{n} \sum_k k \sum_{i \geq k} \Pr(\tau = i)$$
$$= \sum_k (k+1) \left(1 - \frac{k}{n}\right) \Pr(\tau = k) + \frac{1}{n} \sum_i \Pr(\tau = i) \sum_{1 \leq k \leq i} k$$
$$= \mathrm{E}\left((\tau+1)\left(1 - \frac{\tau}{n}\right)\right) + \frac{1}{n} \mathrm{E}\left(\frac{\tau(\tau+1)}{2}\right) = \mathrm{E}(\tau) + 1 - \frac{\mathrm{E}(\tau) + \mathrm{E}(\tau^2)}{2n},$$

which is indeed $\widehat{\mathrm{E}}(\lambda) + \widehat{\mathrm{E}}(\mu)$.

The probability generating function for ρ, $R(z)$, is given by

$$R(z) = \sum_k \left(1 - \frac{k-1}{n}\right) c_{k-1} z^k + \frac{1}{n} \sum_k z^k \sum_{i \geq k} c_i$$
$$= \sum_k \left(1 - \frac{k}{n}\right) c_k z^{k+1} + \frac{1}{n} \sum_i c_i \sum_{1 \leq k \leq i} z^k \qquad (1.21)$$
$$= zC(z) - \frac{z^2 C'(z)}{n} + \frac{z(C(z)-1)}{n(z-1)}.$$

For the higher order factorial moments of ρ first note that

$$R(z) = zC(z) - \frac{z^2}{n} C'(z) + A(z),$$

where $A(z)$ is given by equation (9), so that it is enough to consider the derivatives of

$$D(z) = zC(z) - \frac{z^2}{n} C'(z).$$

We obtain that

$$D(w+1) = (w+1)C(w+1) - \frac{(w+1)^2}{n} C'(w+1)$$
$$= (w+1)(C_0 + C_1 w + C_2 w^2 + \cdots) - \frac{(w+1)^2}{n}(C_1 + 2C_2 w + \cdots);$$

hence

$$D_l = C_l + C_{l-1} - \frac{(l+1)C_{l+1} + 2lC_l + (l-1)C_{l-1}}{n}.$$

From equation 11

$$A_l = \frac{C_l + C_{l+1}}{n},$$

so that

$$R_l = \left(1 - \frac{l-1}{n}\right) C_{l-1} + \left(1 - \frac{2m-1}{n}\right) C_l - \frac{l}{n} C_{l+1}. \tag{1.22}$$

In particular

$$R'(1) = 1 + \left(1 - \frac{1}{n}\right) C'(1) - \frac{1}{2n} C''(1), \tag{1.23}$$

as expected (equations (6) and (13)), and

$$R''(1) = \left(2 - \frac{2}{n}\right) C'(1) + \left(1 - \frac{3}{n}\right) C''(1) - \frac{2}{3n} C'''(1). \tag{1.24}$$

1.8. The covariance of λ and μ

By definition, the covariance of λ and μ is given by

$$\widehat{\mathrm{cov}}(\lambda, \mu) = \hat{\mathrm{E}}(\lambda\mu) - \hat{\mathrm{E}}(\lambda)\,\hat{\mathrm{E}}(\mu),$$

and their correlation is

$$\widehat{\mathrm{cor}}(\lambda, \mu) = \frac{\widehat{\mathrm{cov}}(\lambda, \mu)}{\sqrt{\widehat{\mathrm{var}}(\lambda)\,\widehat{\mathrm{var}}(\mu)}}.$$

It seems that λ and μ must be negatively correlated for permutation invariant weights; if we pick a certain x and a certain f, and if turns out that $\mu(x, f)$ is larger than average, we expect that $\lambda(x, f)$ will be smaller than average because for each value of ρ, if w is permutation invariant, then λ and μ are almost identically distributed. However there are permutation invariant weights such that the correlation of λ and μ is positive; an example is presented in Appendix B.

Our goal in this section is to find the value of $\widehat{\mathrm{cov}}(\lambda, \mu)$ as a function of the moments of r.

We start from

$$\begin{aligned}\widehat{\mathrm{var}}(\rho) &= \hat{\mathrm{E}}(\rho^2 - \hat{\mathrm{E}}(\rho)^2) = \hat{\mathrm{E}}(\lambda^2 + \mu^2 + 2\lambda\mu) - \hat{\mathrm{E}}(\lambda)^2 - \hat{\mathrm{E}}(\mu)^2 - 2\,\hat{\mathrm{E}}(\lambda)\,\hat{\mathrm{E}}(\mu) \\ &= \widehat{\mathrm{var}}(\lambda) + \widehat{\mathrm{var}}(\mu) + 2\,\widehat{\mathrm{cov}}(\lambda, \mu).\end{aligned}$$

Replacing the variance by its expression in terms of derivatives of the probability generating function, we obtain

$$R''(1) + R'(1) - R'(1)^2 = L''(1) + L'(1) - L'(1)^2 + M''(1) + M'(1) - M'(1)^2 + 2\widehat{\mathrm{cov}}(\lambda, \mu).$$

But $R'(1) = L'(1) + M'(1)$, and hence

$$R''(1) - 2L'(1)M'(1) = L''(1) + M''(1) + 2\widehat{\mathrm{cov}}(\lambda, \mu),$$

and finally

$$\widehat{\mathrm{cov}}(\lambda, \mu) = \frac{1}{2}\left(R''(1) - L''(1) - M''(1) - 2L'(1)M'(1)\right). \tag{1.25}$$

After expressing all the factorial moments in terms of the factorial moments of the total number of cyclic elements, r, (equations (5), (12), and (22)), we obtain (with the help of a computer) that

$$\widehat{\mathrm{cov}}(\lambda, \mu) = \frac{1}{12} + \left(\frac{1}{6} - \frac{1}{2n}\right) C'(1) + \left(\frac{1}{6} - \frac{3}{4n}\right) C''(1) - \frac{1}{6n} C'''(1) - \left(\frac{1}{4} - \frac{1}{2n}\right) C'(1)^2 + \frac{1}{4n} C'(1) C''(1). \tag{1.26}$$

1.9. A simple example – permutations

In this section, as a quick check, we shall examine a very simple permutation invariant weight. More intricate problems will be discussed in the following chapters.

Assume that all permutations are equally likely and all other mappings have probability 0. More precisely, the probability weight is defined by

$$w(f) = \begin{cases} 1/n!, & \text{if } f \text{ is a permutation;} \\ 0, & \text{otherwise.} \end{cases}$$

This weight is clearly permutation invariant.

All the elements are always cyclic, hence

$$C(z) = z^n \qquad \mathbf{E}(r) = n \qquad \mathbf{E}(r^2) = n^2 \qquad \mathbf{var}(r) = 0.$$

From Theorems 9 and 10, we obtain that

$$\hat{\mathbf{E}}(\lambda) = \frac{n+1}{2},$$

and

$$\hat{\mathbf{E}}(\mu) = \frac{n+1}{2} - \frac{1}{2n}(n+n^2) = 0,$$

as expected. Equation (8) results in

$$\widehat{\mathrm{var}}(\lambda) = \frac{n^2-1}{12},$$

and equation (26) confirms that $\widehat{\mathrm{cov}}(\lambda,\mu) = 0$. Theorem 8 gives

$$\begin{aligned} L(z) &= \frac{z}{z-1}\int_0^z t^{n-1}\,dt = \frac{z}{z-1}\frac{z^n-1}{n} \\ &= \frac{1}{n}\sum_{k\geq 1} z^k, \end{aligned}$$

and (not too surprisingly)

$$M(z) = 1 + \frac{z(z^n-1)}{n(z-1)} - \frac{z(z^n-1)}{n(z-1)} = 1.$$

For the number of cycles, β we obtain, from equation (15), that the probability generating function is

$$B(z) = \frac{1}{n!}\sum_j \begin{bmatrix} n \\ j \end{bmatrix} z^j.$$

Equations (19) and (20) translate into

$$\mathbf{E}(\beta) = H_n$$

and

$$\widehat{\mathrm{var}}(\beta) = H_n^2 - H_n^{(2)} + H_n - H_n^2 = H_n - H_n^{(2)},$$

a slightly less known fact.

Corollary 5 is the somewhat amazing truth, that no matter how many men will mix up their hats,[3] on average only one of them will get his hat back.

[3] It is not my intention to be sexist, but women never mix their hats.

1.10. Other types of invariance

A weight function w, was called permutation invariant, if for any permutation p and any mapping f, we have $w(f) = w(p \circ f)$. In a similar manner we say that a weight function w is *labelling invariant* if for any permutation p and any mapping f, we have $w(f) = w(f \circ p)$.

Any mapping f induces an equivalence relation over its domain, defined by $x \equiv y$ iff $f(x) = f(y)$. A weight function is labelling invariant if any two mappings that induce the same equivalence relation are equally likely.

A weight function w is *isomorphic invariant* if for any permutation p and any mapping f, we have $w(f) = w(p^{-1} \circ f \circ p)$. In other words, a weight function is isomorphic invariant if any two mappings that have the same (unlabelled) graph are equally likely.

Theorem 17. *Any weight function, w, is equivalent to a isomorphic invariant weight function, w', in the sense that w and w' imply the same distribution for μ, λ, ρ, and τ.*

Proof: Define

$$w' = \frac{1}{n!} \sum_p w(p^{-1} \circ f \circ p),$$

where p ranges over all permutations of $\{1, \ldots, n\}$. ∎

This theorem simplifies the study of the possible probability generating functions for μ, λ, ρ, and τ. For instance, on $F[3]$, the most general probability generating functions for these quantities depend only on 9 parameters (which must sum to 1), corresponding to the probabilities of the 9 isomorphic mappings on 3 elements, and not on 27 parameters corresponding to the 27 mappings in $F[3]$. Such a study shows that the propositions 6, 7, and 16 are independent in the sense that for any subset of them there are (non-invariant) weight functions that satisfy all the propositions in the chosen subset, and do not satisfy the other propositions.

The three types of invariance defined so far (permutation invariant, labelling invariant, and isomorphic invariant) are clearly independent, because there are weight functions that have one property, but do not have the other two. However if any two of the invariance properties are present, all three hold. For instance, permutation invariance and labelling invariance imply isomorphic invariance:

$$w(f) = w(p^{-1} \circ f) = w(p^{-1} \circ f \circ p),$$

and permutation invariance and isomorphic invariance imply labelling invariance:

$$w(f) = w(p \circ f) = w(p^{-1} \circ p \circ f \circ p) = w(f \circ p).$$

We say that a weight function is *strongly invariant* if it is both labelling invariant and permutation invariant. By Theorem 17 and the above observation, any permutation invariant weight function is equivalent (from the point of view of the distribution of μ, λ, ρ, and τ) to a strongly invariant weight function. Similarly, any labelling invariant weight is equivalent to strongly invariant weight, and therefore all the relations between the distributions of μ, λ, ρ, and τ, that hold for permutation invariant weights also hold for labelling invariant weights.

Let $\Delta(f)$ be the multiset $\{d(1), d(2), \ldots, d(n)\}$, where $d(i)$ is the number of elements in $\{1, \ldots, n\}$ where f takes the value i (that is, the indegree of i in the graph of f). With this definition, a weight function w is strongly invariant if for any two mappings f and g that satisfy $\Delta(f) = \Delta(g)$ we have $w(f) = w(g)$.

Clearly for any mapping f, we must have $\sum_{1 \le i \le n} d(i) = n$, so that $\Delta(f)$ is just a partition of n. Therefore a strongly invariant weight is completely characterized by associating to each partition of n a certain probability. For instance if $n = 3$ there are three partitions: $[1,1,1]$, $[2,1]$, and $[3]$. Denoting their probabilities by $w[1,1,1]$, $w[2,1]$, and $w[3]$, it follows that for any strongly invariant weight on $F[3]$, the generating function $C(z)$ must have the form

$$C(z) = \left(w[3] + \frac{1}{3}w[2,1]\right)z + \frac{2}{3}w[2,1]z^2 + w[1,1,1]z^3.$$

Chapter 2

The uniform distribution model

The obvious permutation invariant weight on the space of finite functions, $F[n]$, is the uniform distribution. A considerable number of results are known about this situation; Appendix A contains a bibliography on random mappings that lists over twenty relevant papers. The main results were obtained by Harris [Harris60] and Stepanov [Stepanov69].

In this chapter we shall use the concepts of the first chapter, both to derive some old results in the new setting, and to obtain some new formulæ. The last two sections do not make use of the permutation invariant property of the uniform distribution, but share with the first chapter the use of combinatorics on words as a main tool.

2.1. Preliminaries

The section presents some mathematical entities that will be used later.

2.1.1. The r-Stirling numbers

Stirling numbers of the second kind are denoted by $\left\{ {n \atop m} \right\}$; they are defined combinatorially as the number of partitions of the set $\{1,\ldots,n\}$ into m non-empty disjoint unlabelled sets. Good expositions of their properties can be found, for example, in [Comtet74], [Riordan58], or [Jordan47].

The r-Stirling numbers of the second kind, $\left\{ {n \atop m} \right\}_r$, count certain restricted partitions and are defined, for all integers $r \geq 0$, as the number of partitions of the set $\{1, \ldots, n\}$ into m non-empty disjoint subsets, such that the numbers $1, 2, \ldots, r$ are in distinct subsets.

The properties of the r-Stirling numbers are discussed in [Broder84a]. They were also studied under different names and notations in [Nielsen23], [Carlitz80], and [Koutras82]. Their asymptotics were studied in detail in [IM65]. Here we shall need only the fact that

$$\left\{ {n+m \atop n} \right\}_r = \frac{n^{2m}}{(2m)!!} + O(n^{2m-1}) \tag{2.1}$$

as $n \to \infty$, for fixed m and r. (The notation $x!!$ means $x(x-2)(x-4)\ldots.$)

The r-Stirling numbers satisfy a recurrence similar to the recurrence for Stirling numbers, namely

$$\begin{aligned} \left\{ {n \atop m} \right\}_r &= 0, && n < r, \\ \left\{ {n \atop m} \right\}_r &= \delta_{m,r}, && n = r, \\ \left\{ {n \atop m} \right\}_r &= m\left\{ {n-1 \atop m} \right\}_r + \left\{ {n-1 \atop m-1} \right\}_r, && n > r. \end{aligned} \tag{2.2}$$

2.1.2. The Q-series

Knuth defines the infinite Q-series in [Knuth85] as

$$Q_n(a_1, a_2, \ldots) = \sum_{k \geq 1} \frac{n^{\underline{k}}}{n^k} a_k. \tag{2.3}$$

For any given sequence a_1, a_2, ..., this function depends only on n. In particular, $Q_n(1,1,1,\ldots)$ is denoted $Q(n)$. The asymptotic behavior of $Q(n)$ is well understood ([Ramanujan12], [Knuth73a, §1.2.11.3]):

$$Q(n) = \sqrt{\frac{\pi n}{2}} - \frac{1}{3} + \frac{1}{12}\sqrt{\frac{\pi}{2n}} + \cdots. \tag{2.4}$$

The Q-series are relevant to many problems in the analysis of algorithms [Knuth85], for instance the representation of equivalence relations [KS78], hashing

[Knuth73b, §6.4], interleaved memory [KR75], labelled tree enumeration [Moon71], optimal caching [Knuth85], permutations *in situ* [Stanford81], and random mappings [Knuth81, §3.1].

It can readily be shown that the Q-series satisfy the recurrence

$$Q_n(a_1, 2a_2, 3a_3, \ldots) = nQ_n(a_1, a_2 - a_1, a_3 - a_2, \ldots). \tag{2.5}$$

From this recurrence it follows that

$$\begin{aligned} Q_n(1, 2, 3, \ldots) &= n; \\ Q_n(1^2, 2^2, 3^2, \ldots) &= nQ(n); \\ Q_n(1^3, 2^3, 3^3, \ldots) &= 2n^2 - nQ(n). \end{aligned} \tag{2.6}$$

In general there exist integral and positive coefficients $q_{m,k}$ such that

$$Q_n(1^m, 2^m, 3^m, \ldots) = q_{m,0}Q_n^{[m]} - q_{m,1}Q_n^{[m-1]} + q_{m,2}Q_n^{[m-2]} - \cdots, \tag{2.7}$$

where

$$Q_n^{[m]} = \begin{cases} n^{(m+1)/2}, & \text{if } m \text{ is odd;} \\ n^{m/2}Q(n), & \text{if } m \text{ is even} \end{cases}$$

The leading coefficient has a simple expression, $q_{m,0} = (m-1)!!$ (see [Knuth85] for details). Consequently for s fixed,

$$Q_n(1^{2s}, 2^{2s}, \ldots) = \sum_{k \geq 1} \frac{n^{\underline{k}}}{n^k} k^{2s} = (2s-1)!!\, n^s Q(n) + O(n^s), \tag{2.8}$$

and

$$Q_n(1^{2s+1}, 2^{2s+1}, \ldots) = \sum_{k \geq 1} \frac{n^{\underline{k}}}{n^k} k^{2s+1} = (2s)!!\, n^{s+1} + O(n^{s+1/2}). \tag{2.9}$$

There is an interesting relation between Q-series and r-Stirling numbers: For all $h \geq 1$ we have

$$Q_n\left(\left\{{h \atop 1}\right\}_r, 2\left\{{h+1 \atop 2}\right\}_r, \ldots\right) = \sum_{k \geq 1} \frac{n^{\underline{k}}}{n^k} k \left\{{h+k-1 \atop k}\right\}_r = n^{\underline{h}} \frac{n^{\underline{r}}}{n^r}, \tag{2.10}$$

and in particular for $r = 0$ (i.e., regular Stirling numbers) we have

$$Q_n\left(\left\{{h \atop 1}\right\}, 2\left\{{h+1 \atop 2}\right\}, \ldots\right) = n^{\underline{h}}. \tag{2.11}$$

Both this equations are an immediate consequence of the recurrences (2) and (5).

Another noteworthy particularization of equation (5) is

$$\sum_{k\geq 1}\frac{n^{\underline{k}}}{n^k}ka_k = n\Big(\sum_{k>i}\frac{n^{\underline{k}}}{n^k}(a_k - a_{k-1}) + \frac{n^{\underline{i}}}{n^i}a_i\Big). \tag{2.12}$$

2.1.3. Asymptotics of certain Q-series

In this section we present a fairly general method for obtaining the leading term in the asymptotic expansion of Q-series of the form $\sum_{k\geq 1} f(k)n^{\underline{k}}/n^k$ where f is a differentiable function on the interval $[1,\infty)$, and where $f'(x) = O(x^\alpha)$ for some constant α and all $x \in [1,\infty)$. Clearly

$$f(x) = f(1) + \int_1^x f'(t)\,dt,$$

so this condition implies that

$$f(x) = \begin{cases} O(x^{\alpha+1}), & \text{if } \alpha > -1; \\ O(\ln x), & \text{if } \alpha = -1; \\ O(1), & \text{if } \alpha < -1. \end{cases} \tag{2.13}$$

We start by noticing that

$$\frac{n^{\underline{k}}}{n^k} = \prod_{1\leq i<k}\Big(1 - \frac{i}{n}\Big) \leq \prod_{1\leq i<k} e^{-i/n} = e^{-k(k-1)/(2n)}, \tag{2.14}$$

and hence $f(k)n^{\underline{k}}/n^k$ is exponentially small for $k \geq n^{1/2+\epsilon}$, for any $\epsilon > 0$.

For $k \leq n^{1/2+\epsilon}$ we can use the Stirling expansion to obtain

$$\begin{aligned} \ln\frac{n!}{(n-k)!\,n^k} &= (n+\tfrac{1}{2})\ln n - n - (n-k+\tfrac{1}{2})\ln(n-k) \\ &\quad + (n-k) - k\ln n + O(n^{-1}) \\ &= (n-k+\tfrac{1}{2})\ln\frac{n}{n-k} - k + O(n^{-1}). \end{aligned} \tag{2.15}$$

Expanding the logarithm in its Taylor series we get

$$\ln\frac{n}{n-k} = \ln\Big(1 + \frac{k}{n-k}\Big) = \frac{k}{n-k} - \frac{k^2}{2(n-k)^2} + O\Big(\frac{k^3}{(n-k)^3}\Big), \tag{2.16}$$

which combined with equation (15) leads, when $k \le n^{1/2+\epsilon}$, to

$$\ln \frac{n!}{(n-k)!\,n^k} = \frac{k}{2(n-k)} - \frac{k^2}{2(n-k)} + O\left(\frac{k^3}{(n-k)^2}\right) = -\frac{k^2}{2n} + O(n^{-1/2+3\epsilon}). \tag{2.17}$$

Therefore

$$\frac{n^{\underline{k}}}{n^k} = e^{-k^2/(2n)}\big(1 + O(n^{-1/2+3\epsilon})\big), \qquad k \le n^{1/2+\epsilon}. \tag{2.18}$$

These results help us to prove the following

Lemma 1. *Let f be a differentiable function on $[1,\infty)$ such that $f'(x) = O(x^\alpha)$ on $[1,\infty)$. Then*

$$\sum_{k\ge 1} \frac{n^{\underline{k}}}{n^k} f(k) = \sum_{k\ge 1} e^{-k^2/(2n)} f(k)\big(1 + O(n^{-1/2+\epsilon})\big),$$

for any $\epsilon > 0$. (The constant implied by O does not depend on ϵ.)

Proof: This follows from equation (18) with a proper choice of ϵ, and the fact the the sum of the terms corresponding to $k \ge n^{1/2+\epsilon/3}$ is exponentially small. ∎

The next step in our approximation is to convert the sum to an integral using Euler's summation formula. To evaluate the error term, we shall need the following

Lemma 2. *As $n \to \infty$,*

$$\int_1^\infty e^{-x^2/n} x^\beta\, dx = \begin{cases} n^{(\beta+1)/2}\Gamma\big((\beta+1)/2\big)/2 + O(n^{-(\beta+1)/2}), & \text{if } \beta > -1; \\ (\ln n)/2 - \gamma/2 + O\big(\ln(n)/n\big), & \text{if } \beta = -1; \\ O(1), & \text{if } \beta < -1. \end{cases}$$

Proof: For $\beta > -1$ by making the substitution $y \leftarrow x^2/n$ we obtain

$$\begin{aligned} \int_1^\infty e^{-x^2/n} x^\beta\, dx &= \frac{n^{(\beta+1)/2}}{2} \int_{1/n}^\infty e^{-y} y^{(\beta-1)/2}\, dy \\ &= \frac{n^{(\beta+1)/2}}{2}\Gamma\big((\beta+1)/2\big) - \int_0^{1/n} e^{-y} y^{(\beta-1)/2}\, dy, \end{aligned}$$

but

$$\int_0^{1/n} e^{-y} y^{(\beta-1)/2}\, dy \le \int_0^{1/n} y^{(\beta-1)/2}\, dy = \frac{n^{-(\beta+1)/2}}{(\beta+1)/2} = O(n^{-(\beta+1)/2}),$$

which proves the first case of the lemma.

For the case $\beta = -1$, we use the same substitution, $y \leftarrow x^2/n$, and integration by parts, to get

$$\int_1^\infty e^{-x^2/n} x^{-1}\,dx = \frac{1}{2}\int_{1/n}^\infty \frac{e^{-y}}{y}\,dy = \frac{e^{-y}\ln y}{2}\Big|_{1/n}^\infty + \frac{1}{2}\int_{1/n}^\infty e^{-y}\ln y\,dy$$
$$= \frac{e^{-1/n}\ln n}{2} + \frac{1}{2}\int_0^\infty e^{-y}\ln y\,dy - \frac{1}{2}\int_0^{1/n} e^{-y}\ln y\,dy.$$

But it is known that

$$\int_0^\infty e^{-y}\ln y\,dy = -\gamma,$$

where γ is Euler's constant (0.5772 ...), and

$$\left|\int_0^{1/n} e^{-y}\ln y\,dy\right| \le \left|\int_0^{1/n} \ln y\,dy\right| = \frac{\ln n + 1}{n} = O\left(\frac{\ln n}{n}\right).$$

Hence

$$\int_1^\infty e^{-x^2/n} x^{-1}\,dx = \frac{e^{-1/n}\ln n}{2} - \frac{\gamma}{2} + O\left(\frac{\ln n}{n}\right)$$
$$= \frac{\ln n}{2} - \frac{\gamma}{2} + O\left(\frac{\ln n}{n}\right).$$

Finally, if $\beta < -1$, then

$$\int_1^\infty e^{-x^2/n} x^\beta\,dx \le e^{-1/n}\int_1^\infty x^\beta\,dx = -\frac{e^{-1/n}}{\beta+1} = O(1).$$

∎

We can now prove the main result of this section.

Theorem 3. *Let f be a differentiable function on $[1,\infty)$ such that $f'(x) = O(x^\alpha)$ on $[1,\infty)$. Then, for any positive ϵ, as $n \to \infty$,*

$$\sum_{k\ge 1} \frac{n^{\underline{k}}}{n^k} f(k) = \left(\int_1^\infty e^{-x^2/(2n)} f(x)\,dx\right)\left(1 + O(n^{-1/2+\epsilon})\right) + O(g(n)),$$

where

$$g(n) = \begin{cases} n^{(\alpha+1)/2}, & \text{if } \alpha > -1; \\ \ln n, & \text{if } \alpha = -1; \\ 1, & \text{if } \alpha < -1. \end{cases}$$

Proof: By Euler's summation formula, if h is differentiable, then

$$\sum_{k\geq 1} h(k) = \int_1^\infty h(x)\,dx - \frac{1}{2}h(x)\Big|_1^\infty + \int_1^\infty (x - \lfloor x\rfloor - 1/2)h'(x)\,dx$$
$$= \int_1^\infty h(x)\,dx - \frac{1}{2}h(x)\Big|_1^\infty + O\left(\int_1^\infty |h'(x)|\,dx\right).$$

If we set $h(x) \leftarrow e^{-x^2/(2n)}f(x)$ then as $n \to \infty$,

$$\frac{1}{2}h(x)\Big|_1^\infty = O(1),$$

and

$$h'(x) = \left(-\frac{x}{n}f(x) + f'(x)\right)e^{-x^2/(2n)}.$$

Hence

$$\sum_{1\leq k} e^{-k^2/(2n)}f(k) = \int_1^\infty e^{-x^2/(2n)}f(x)\,dx + O\left(\frac{1}{n}\int_1^\infty \left|e^{-x^2/(2n)}xf(x)\right|\,dx\right)$$
$$+ O\left(\int_1^\infty \left|e^{-x^2/(2n)}f'(x)\right|\,dx\right) + O(1).$$

We need now to evaluate on a case by case basis the integrals

$$A = \frac{1}{n}\int_1^\infty e^{-x^2/(2n)}\,|xf(x)|\,dx,$$

and

$$B = \int_1^\infty e^{-x^2/(2n)}\,|f'(x)|\,dx.$$

If $\alpha > -1$ then $xf(x) = O(x^{\alpha+2})$, and by the first case of Lemma 2 we have $A = O(n^{(\alpha+3)/2})/n = O(n^{(\alpha+1)/2})$ and $B = O(n^{(\alpha+1)/2})$.

If $\alpha = -1$, then $xf(x) = O(x\ln x)$ and

$$A = O\left(\frac{1}{n}\int_1^\infty e^{-x^2/(2n)}x\ln x\,dx\right);$$

Integrating by parts and using the second case of Lemma 2, we obtain

$$\int_1^\infty e^{-x^2/(2n)}x\ln x\,dx = n\int_1^\infty e^{-x^2/(2n)}x^{-1}\,dx = O(n\ln n),$$

so that $A = O(\ln n)$. Also, directly from the second case of Lemma 2, $B = O(\ln n)$.

If $\alpha < -1$ then $xf(x) = O(x)$ and by the first case of Lemma 2, $A = O(n)/n = O(1)$. By the third case, $B = O(1)$.

We conclude that

$$\sum_{k\geq 1} e^{-k^2/(2n)} f(k) = \int_1^\infty e^{-x^2/(2n)} f(x)\,dx + O(g(n)),$$

where $g(n)$ has the desired form, and the theorem is proved by applying Lemma 1 to the last equation. ∎

As an example we can compute, for $s > 0$, the Q-series $\sum_{k\geq 1} k^s n^{\underline{k}}/n^k$. In this case $\alpha = s - 1$ and we obtain using the first case of Lemma 2

$$\begin{aligned}\sum_{k\geq 1} \frac{n^{\underline{k}}}{n^k} k^s &= \left(\int_1^\infty e^{-x^2/(2n)} x^s\,dx\right)\left(1 + O(n^{-1/2+\epsilon})\right) + O(n^{s/2}) \\ &= \frac{(2n)^{(s+1)/2}}{2}\Gamma\left((s+1)/2\right) + O(n^{s/2+\epsilon}).\end{aligned}$$

In particular if s is an odd integer, the result is

$$2^{(s-1)/2} n^{(s+1)/2}((s-1)/2)! + O(n^{s/2+\epsilon}) = (s-1)!!\, n^{(s+1)/2} + O(n^{s/2+\epsilon}),$$

and if s is an even integer, the result is

$$\begin{aligned}&\frac{(2n)^{(s+1)/2}}{2}\frac{s-1}{2}\frac{s-3}{2}\cdots\frac{1}{2}\Gamma\left(\frac{1}{2}\right) + O(n^{s/2+\epsilon}) \\ &\qquad = (s-1)!!\, n^{(s+1)/2}\sqrt{\frac{\pi}{2}} + O(n^{s/2+\epsilon}),\end{aligned}$$

in agreement with equations (8) and (9).

2.1.4. Abelian identities

Sums of the type

$$A_n(x,y;p,q) = \sum_k \binom{n}{k}(x+k)^{k+p}(y+n-k)^{n-k+q}, \qquad p,q,n \text{ integers},$$

are called "Abelian binomial sums" by Riordan [Riordan68], [Riordan69].

With this notation, the famous Abel identity [Abel1826] becomes

$$A_n(x,y;-1,0) = \frac{(x+y+n)^n}{x}. \tag{2.19}$$

This is sometimes written as an identity in three variables

$$\sum_k \binom{n}{k}(x+kz)^{k-1}(y+(n-k)z)^{n-k} = \frac{(x+y+nz)^n}{x}, \tag{2.20}$$

via the substitutions $x \leftarrow x/z$, and $y \leftarrow y/z$. Equation (19) can also be written as

$$n^n + \sum_{k\geq 1} \binom{n}{k} x(x+k)^{k-1}(n-k)^{n-k} = (x+n)^n,$$

which, after taking derivatives with respect to z and setting $x \leftarrow 0$, becomes

$$\sum_{k\geq 0} \binom{n}{k} k^{k-1}(n-k)^{n-k} = n^n. \tag{2.21}$$

Another well known example of an identity involving Abelian sums is the Cauchy formula [Cauchy1826]

$$A_n(x,y;0,0) = \sum_k \binom{n}{k}(x+k)^k(y+n-k)^{n-k} = \sum_k \binom{n}{k} k!(x+y+n)^{n-k}, \tag{2.22}$$

which for $x = y = 0$ results in

$$\sum_k \binom{n}{k} k^k(n-k)^{n-k} = \sum_k \binom{n}{k} k! n^{n-k} = n^n(Q(n)+1). \tag{2.23}$$

Riordan found a recurrence and a symmetry formula for A_n and used them to prove these identities and also to derive similar "Abelian identities" iteratively for p and q between -3 and 3. Another proof method, due to Françon [Françon74], is based on the FF-encoding applied to a suitably chosen family of mappings. In this manner Françon proved the Abel identity and the Cauchy formula by counting arguments. The author obtained a general explicit expression for the Abelian identities ([Broder83]), for all $p, q \geq 0$, using similar word counting arguments. For $x = y = 0$ the general identity is

$$\sum_k \binom{n}{k} k^{k+p}(n-k)^{n-k+q} = n^{n-1} \sum_{k,l\geq 0} \frac{n^{k+l}}{n^{k+l}} \left\{ {k+p \atop k} \right\} \left\{ {l+q \atop l} \right\} (k+l). \tag{2.24}$$

In particular

$$\sum_k \binom{n}{k} k^{k+p}(n-k)^{n-k} = n^n \sum_{k\geq 0} \frac{n^{\underline{k}}}{n^k} \left\{ {k+p \atop k} \right\}, \tag{2.25}$$

since

$$\begin{aligned}\sum_k \binom{n}{k} k^{k+p}(n-k)^{n-k} &= n^{n-1} \sum_{k,l\geq 0} \frac{n^{\underline{k+l}}}{n^{k+l}} \left\{ {k+p \atop k} \right\} (k+l) = n^{n-1} \sum_{k\geq 1} \frac{n^{\underline{k}}}{n^k} k \sum_{0\leq l\leq k} \left\{ {l+p \atop l} \right\} \\ &= n^n \left(\left\{ {0+p \atop 0} \right\} + \sum_{k\geq 1} \frac{n^{\underline{k}}}{n^k} \left\{ {k+p \atop k} \right\} \right) = n^n \sum_{k\geq 0} \frac{n^{\underline{k}}}{n^k} \left\{ {k+p \atop k} \right\}.\end{aligned}$$

(The penultimate step is based on equation (5).)

Another Abelian type identity that we shall need is

$$\sum_k \binom{n}{k} k^{k+p} \frac{k^{\underline{i}}}{k^i} (n-k)^{n-k} = n^n \sum_k \frac{n^{\underline{k}}}{n^k} \left\{ {k+p \atop k} \right\}_i, \qquad i > 0. \tag{2.26}$$

(This and similar identities are proved in [Broder84b].)

2.2. The distribution of the number of cyclic elements

We have seen that the probability distribution of many of the random variables of interest is closely related to the distribution of the number of cyclic elements, τ; hence, we start by computing this distribution.

Lemma 4. *Given a uniform probability distribution $F[n]$,*

$$\mathbf{Pr}(\tau = k) = \frac{k n^{\underline{k}}}{n^{k+1}}.$$

Proof: Recall that the repetition index ν of a mapping f is the maximum number i such that $f(1), f(2), \ldots, f(i)$ are distinct. Clearly, if all functions are equally likely, then

$$\mathbf{Pr}(\nu = k) = \frac{n(n-1)\ldots(n-k+1)k}{n^{k+1}},$$

and by Theorem 1.13 this is the same as the probability that $\tau = k$. ∎

We immediately obtain

Theorem 5. *Given a uniform probability distribution $F[n]$, the probability generating function of the number of cyclic elements is*

$$C(z) = \frac{1}{n}\sum_{k\geq 1}\frac{n^{\underline{k}}}{n^k}kz^k.$$

∎

Note that (equation (6)) we indeed have

$$\sum_k \Pr(\tau = k) = \frac{Q_n(1,2,\ldots)}{n} = 1. \tag{2.27}$$

The generating function $C(z)$ has a nice form as a Q-series:

$$\begin{aligned} C(z) &= \frac{1}{n}Q_n(z,2z^2,3z^3,\ldots) = \frac{1}{n}Q_n(z,z^2-z,z^3-z^2,\ldots) \\ &= (z-1)Q_n(1,z,z^2,\ldots)+1. \end{aligned} \tag{2.28}$$

The factorial moments of τ can be expressed in terms of the Q-series as

$$\begin{aligned} C^{(l)} &= \frac{1}{n}\sum_{k\geq 1}\frac{n^{\underline{k}}}{n^k}k\underline{k^l} = \frac{1}{n}\sum_i \begin{bmatrix} l \\ i \end{bmatrix}(-1)^{l-i}\sum_{k\geq 1}\frac{n^{\underline{k}}}{n^k}k^{i+1} \\ &= \frac{1}{n}\sum_i \begin{bmatrix} l \\ i \end{bmatrix}(-1)^{l-i}Q_n(1,2^{i+1},3^{i+1},\ldots). \end{aligned} \tag{2.29}$$

The same identity can be derived as follows. Recall that

$$C(w+1) = C_0 + wC_1 + w^2C_2 + \cdots,$$

where $C_l = C^{(l)}/l!$. From equation (28) we have

$$C(w+1) = wQ_n(1,w+1,(w+1)^2\ldots)+1,$$

and therefore

$$\begin{aligned} C_l &= Q_n\left(\binom{0}{l-1},\binom{1}{l-1},\binom{2}{l-1},\ldots\right) \\ &= \frac{1}{n}Q_n\left(\binom{1}{l},2\binom{2}{l},3\binom{3}{l},\ldots\right), \end{aligned} \tag{2.30}$$

from which equation (29) follows immediately.

For fixed l, using the asymptotics of the Q-series (equations (7), (8), and (9)), we obtain

$$\begin{aligned} C^{(l)} &= \frac{1}{n}\sum_i \begin{bmatrix} l \\ i \end{bmatrix} (-1)^{l-1} i!! \, Q_n^{[i+1]}\left(1 + O(n^{-1/2})\right) \\ &= \frac{l!!}{n} Q_n^{[l+1]}\left(1 + O(n^{-1/2})\right). \end{aligned} \tag{2.31}$$

In particular

$$\begin{aligned} C'(1) &= Q(n), \\ C''(1) &= 2n - 2Q(n), \\ C'''(1) &= 3nQ(n) - 9n + 6Q(n). \end{aligned} \tag{2.32}$$

From here

$$\mathbf{E}(r) = C'(1) = Q(n), \tag{2.33}$$

and

$$\begin{aligned} \mathbf{var}(r) &= C''(1) + C'(1) - \left(C'(1)\right)^2 = 2n - Q(n) - Q(n)^2 \\ &= \frac{(4-2\pi)n}{4} - \frac{1}{3}\sqrt{\frac{\pi n}{2}} + O(1). \end{aligned} \tag{2.34}$$

2.3. The distribution of λ and μ

Theorem 6. *Given a uniform probability distribution on $F[n]$, the length of the period and the length of the tail from a starting point chosen uniformly at random satisfy*

$$\widehat{\mathbf{Pr}}(\lambda = i) = \frac{1}{n}\sum_{k \geq i} \frac{n^{\underline{k}}}{n^k}, \qquad 1 \leq i \leq n;$$

$$\widehat{\mathbf{Pr}}(\mu = i) = \frac{1}{n}\sum_{k > i} \frac{n^{\underline{k}}}{n^k}, \qquad 0 \leq i \leq n.$$

Proof: The first relation follows immediately from Lemma 4 and Lemma 1.6; for the second we use Corollary 1.7 and equation (12) to obtain

$$\widehat{\mathbf{Pr}}(\mu = i) = \frac{1}{n}\sum_{k \geq i} \frac{n^{\underline{k}}}{n^k}\left(1 - \frac{k}{n}\right) = \frac{1}{n}\sum_{k > i} \frac{n^{\underline{k}}}{n^k}, \qquad 1 \leq i \leq n;$$

the case $i = 0$ follows from Theorem 1.8. ∎

Corollary 7. *Given a uniform probability distribution on $F[n]$, the probability generating functions for length of the period and the length of the tail from a starting point chosen uniformly at random satisfy*

$$L(z) = zM(z).$$

∎

The factorial moments of λ were computed via equation (1.5), using a computerized formal manipulation system (MAPLE). The first few are listed below.

$$\begin{aligned} L'(1) &= \frac{Q(n)+1}{2}, \\ L''(1) &= \frac{2n - Q(n) - 1}{3} \\ L'''(1) &= \frac{3nQ(n) - 7n + 2Q(n) + 2}{4}. \end{aligned} \tag{2.35}$$

From here

$$\widehat{\mathrm{E}}(\lambda) = \frac{1}{2}\sqrt{\frac{\pi n}{2}} + \frac{1}{3} + \frac{1}{24}\sqrt{\frac{\pi}{2n}} + O(n^{-1}), \tag{2.36}$$

and

$$\begin{aligned} \widehat{\mathrm{var}}(\lambda) &= \frac{8n - 3Q(n)^2 - 4Q(n) - 1}{12} \\ &= \frac{(16-3\pi)n}{24} - \frac{1}{6}\sqrt{\frac{\pi n}{2}} - \frac{\pi}{48} + O(n^{-1/2}). \end{aligned} \tag{2.37}$$

Since $L(z) = zM(z)$, the distribution of μ is a shift of the distribution of λ; we have

$$\widehat{\mathrm{E}}(\mu) = \widehat{\mathrm{E}}(\lambda) - 1 = \frac{1}{2}\sqrt{\frac{\pi n}{2}} - \frac{2}{3} + \frac{1}{24}\sqrt{\frac{\pi}{2n}} + O(n^{-1}), \tag{2.38}$$

and

$$\widehat{\mathrm{var}}(\mu) = \widehat{\mathrm{var}}(\lambda) = \frac{(16-3\pi)n}{24} - \frac{1}{6}\sqrt{\frac{\pi n}{2}} - \frac{\pi}{48} + O(n^{-1/2}), \tag{2.39}$$

and the higher order central moments are also equal.

2.4. The distribution of ρ

In the case of a uniform distribution on $F[n]$, the distribution of ρ is easy to compute by direct arguments. Consider the sequence $x, f(x), f^2(x), \ldots$, for some fixed x. The probability of k distinct values in this sequence is clearly

$$\widehat{\Pr}(\rho = k) = \frac{n(n-1)\ldots(n-k+1)k}{n^{k+1}} = \frac{kn^{\underline{k}}}{n^{k+1}}, \tag{2.40}$$

because all possible sequences are equally likely.

Of course, we can also compute it from Theorem 1.16. We obtain

$$\widehat{\Pr}(\rho = k) = \left(1 - \frac{k-1}{n}\right) \frac{(k-1)n^{\underline{k-1}}}{n^k} + \frac{1}{n^2} \sum_{i \geq k} \frac{n^{\underline{i}}}{n^i} i.$$

Using formula (12) for the sum, and expressing falling powers as binomial coefficients we get

$$\begin{aligned}
\widehat{\Pr}(\rho = k) &= \frac{n-k+1}{n} \binom{n}{k-1} \frac{(k-1)!\,(k-1)}{n^k} + \frac{1}{n} \binom{n}{k} \frac{k!}{n^k} \\
&= \frac{1}{n^{k+1}} \left((n-k+1) \binom{n}{n-k+1} (k-1)!\,(k-1) + \binom{n}{k} k! \right) \\
&= \frac{1}{n^{k+1}} \left(\binom{n}{k} k!\,(k-1) + \binom{n}{k} k! \right) = \frac{k n^{\underline{k}}}{n^{k+1}}.
\end{aligned}$$

If we compare this result with Lemma 4 we see that for a uniform distribution on $F[n]$ the distributions of r and of ρ are identical. Hence, from equations (33) and (34), we have

$$\widehat{\mathrm{E}}(\rho) = Q(n), \tag{2.41}$$

and

$$\widehat{\mathrm{var}}(\rho) = 2n - Q(n)^2 - Q(n). \tag{2.42}$$

From equation (1.26) it follows that

$$\begin{aligned}
\widehat{\mathrm{cov}}(\lambda, \mu) &= \frac{4n - 3Q(n)^2 - 2Q(n) + 1}{12} \\
&= \frac{(8 - 3\pi)n}{24} + \frac{48 - 9\pi}{432} + O(n^{-1/2}),
\end{aligned} \tag{2.43}$$

and

$$\begin{aligned}
\widehat{\mathrm{cor}}(\lambda, \mu) &= \frac{4n - 3Q(n)^2 - 2Q(n) + 1}{8n - 3Q(n)^2 - 4Q(n) - 1} \\
&= \frac{3\pi - 8}{3\pi - 16} - \frac{4(3\pi - 8)}{(3\pi - 16)^2} \sqrt{\frac{\pi}{2n}} + O(n^{-1}) \\
&\approx -0.21668895\ldots.
\end{aligned} \tag{2.44}$$

This is quite a strong negative correlation. It means that if we chose a function f and a point x, both uniformly at random, then if $\lambda(x, f)$ is large, it is very probable that $\mu(x, f)$ is small.

2.5. The distribution of the number of cycles

The probability generating function for the number of cycles, β, is

$$B(z) = \frac{1}{n}\sum_i z^i \sum_k \binom{n}{k}\begin{bmatrix} k \\ i \end{bmatrix} k n^{-k}, \tag{2.45}$$

from Lemma 4 and equations (1.15). Its first factorial moments are obtained from equations (1.19) and (1.20), via formula (5):

$$B'(1) = \frac{1}{n}\sum_{k\geq 1}\frac{n^{\underline{k}}}{n^k} k H_k = \sum_{k\geq 1}\frac{n^{\underline{k}}}{n^k}\frac{1}{k} \tag{2.46}$$

and

$$\begin{aligned} B''(1) &= \frac{1}{n}\sum_{k\geq 1}\frac{n^{\underline{k}}}{n^k} k\left(H_k^2 - H_k''(1)\right) \\ &= \sum_{k\geq 1}\frac{n^{\underline{k}}}{n^k}\left((H_k - H_{k-1})(H_k + H_{k-1}) - \frac{1}{k^2}\right) = \sum_{k\geq 1}\frac{n^{\underline{k}}}{n^k}\frac{2H_{k-1}}{k}. \end{aligned} \tag{2.47}$$

The probability that a function chosen uniformly at random over $F[n]$ has a connected graph is (Corollary 1.12)

$$\mathbf{Pr}(\beta = 1) = \frac{1}{n}\sum_{k\geq 1}\frac{n^{\underline{k}}}{n^k} = \frac{Q(n)}{n}. \tag{2.48}$$

This is one of the earliest results about random mappings, due to Katz [Katz55].

To get the asymptotics of $B'(1)$ (i.e., the expected number of cycles) we use Theorem 3 with $\alpha = -2$ and Lemma 2 with $\beta = -1$. We obtain

$$\begin{aligned} B'(1) = \sum_{k\geq 1}\frac{n^{\underline{k}}}{n^k}\frac{1}{k} &= \left(\int_1^\infty \frac{e^{-x^2/(2n)}}{x}\,dx\right)\left(1 + O(n^{-1/2+\epsilon})\right) + O(1) \\ &= \left(\ln(2n)/2 + O(1)\right)\left(1 + O(n^{-1/2+\epsilon})\right) + O(1) = \frac{\ln n}{2} + O(1). \end{aligned} \tag{2.49}$$

The average number of cycles in a random mapping was first computed (by very different methods) by A. Kruskal [Kruskal54]. He obtained the more precise estimate

$$B'(1) = \frac{\ln 2n}{2} + \frac{\gamma}{2} + o(1). \tag{2.50}$$

The asymptotics of $B''(1)$ are somewhat more difficult. We first replace H_{k-1} in equation (47) by $\ln k + \gamma + O(1/k)$, and then use equation (49) and Theorem 3 to obtain

$$\begin{aligned} B''(1) &= 2\sum_{k\ge1}\frac{n^{\underline{k}}}{n^k}\frac{\ln k}{k} + 2\gamma\sum_{k\ge1}\frac{n^{\underline{k}}}{n^k}\frac{1}{k} + O\Big(\sum_{k\ge1}\frac{n^{\underline{k}}}{n^k}\frac{1}{k^2}\Big) \\ &= 2\sum_{k\ge1}\frac{n^{\underline{k}}}{n^k}\frac{\ln k}{k} + \gamma\ln n + O(1). \end{aligned} \tag{2.51}$$

Because $(\ln x/x)' = (1-\ln x)/x^2$, we can use Theorem 3 with, say, $\alpha = -1.99$, to get

$$\sum_{k\ge1}\frac{n^{\underline{k}}}{n^k}\frac{\ln k}{k} = \Big(\int_1^\infty e^{-x^2/(2n)}\frac{\ln x}{x}\,dx\Big)\big(1+O(n^{-1/2+\epsilon})\big) + O(1). \tag{2.52}$$

With the substitution $y \leftarrow x^2/(2n)$ the integral becomes

$$\begin{aligned} \int_1^\infty e^{-x^2/(2n)}\frac{\ln x}{x}\,dx &= \frac12\int_{1/(2n)}^\infty e^{-y}\frac{\ln\sqrt{2ny}}{y}\,dy \\ &= \frac{\ln 2n}{4}\int_{1/(2n)}^\infty \frac{e^{-y}}{y}\,dy + \frac14\int_{1/(2n)}^\infty e^{-y}\frac{\ln y}{y}\,dy. \end{aligned} \tag{2.53}$$

We already computed the first integral in the proof of Lemma 2. We obtained

$$\int_{1/(2n)}^\infty \frac{e^{-y}}{y}\,dy = \ln 2n - \gamma + O\Big(\frac{\ln n}{n}\Big). \tag{2.54}$$

We can integrate the second integral by parts:

$$\begin{aligned} \int_{1/(2n)}^\infty e^{-y}\frac{\ln y}{y}\,dy &= \frac{e^{-y}(\ln y)^2}{2}\Big|_{1/(2n)}^\infty + \frac12\int_{1/(2n)}^\infty e^{-y}(\ln y)^2\,dy \\ &= -\frac{(\ln 2n)^2}{2} + O(1) + \frac12\int_{\frac{1}{2n}}^\infty e^{-y}(\ln y)^2\,dy \\ &= -\frac{(\ln 2n)^2}{2} + O(1), \end{aligned} \tag{2.55}$$

because

$$\begin{aligned} 0 \le \int_{1/(2n)}^\infty e^{-y}(\ln y)^2\,dy &\le \int_0^\infty e^{-y}(\ln y)^2\,dy \\ &\le \int_1^\infty e^{-y}(\ln y)^2\,dy + \int_0^1 (\ln y)^2\,dy = O(1). \end{aligned}$$

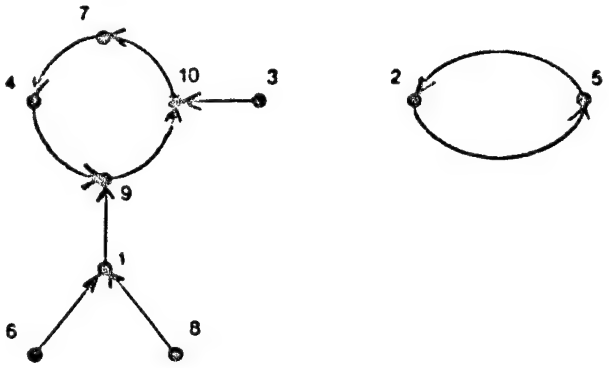


Figure 2.1. An encoding example.

It is obvious that the encoding is 1–1. Hence it suffices to count how many possible legal words exist. First we remark that

1. The length of each word is exactly $s + k$.
2. Within the first $s + k - 1$ positions each of the k letters appears at least once, and the first s positions contain the letters $\overline{a_1}, \ldots, \overline{a_s}$ in that order.

To construct a legal word:

- Partition the first $k + s - 1$ positions into k non-empty subsets such that positions $1, \ldots, s$ are in different subsets. (Each subset corresponds to a certain letter.) This can be done in $\left\{ {k+s-1 \atop k} \right\}_s$ ways.
- Associate to the subsets containing the positions $1, \ldots, s$ the letters $\overline{a_1}, \ldots, \overline{a_s}$ in this order. To each of the remaining $k - s$ subsets associate one of the remaining $k - s$ letters. $((k - s)!$ ways)
- Choose any letter for the last position. (k possibilities)

From this construction it follows that

$$|F_k(S,k)| = (k-s)!\, k \left\{ {k+s-1 \atop k} \right\}_s, \tag{2.59}$$

and therefore, by equation (58),

$$|F_n(S,k)| = \binom{n-s}{k-s} (k-s)!\, n^{n-k} k \left\{ {k+s-1 \atop k} \right\}_s, \tag{2.60}$$

and finally

$$\Pr(\kappa(s)=k)=\frac{|F_n(S,k)|}{n^n}=\frac{1}{n^{\underline{s}}}\frac{n^{\underline{k}}}{n^k}k\left\{{k+s-1\atop k}\right\}_s. \tag{2.61}$$

As a quick check, note that by equation (10), we indeed have

$$\sum_k p_n(S,k)=\frac{1}{n^{\underline{s}}}\sum_k\frac{n^{\underline{k}}}{n^k}k\left\{{k+s-1\atop k}\right\}_s=1. \tag{2.62}$$

Of more interest is the expected value of the size of the transitive closure, that is the sum

$$\mathrm{E}(\kappa(s))=\frac{1}{n^{\underline{s}}}\sum_k\frac{n^{\underline{k}}}{n^k}k^2\left\{{k+s-1\atop k}\right\}_s. \tag{2.63}$$

For fixed s, combining equations (1), (8), and (9), we obtain the estimate

$$\begin{aligned}\frac{1}{n^{\underline{s}}}\sum_k\frac{n^{\underline{k}}}{n^k}k^2\left\{{k+s-1\atop k}\right\}_s&=\frac{1}{n^{\underline{s}}}\sum_k\frac{n^{\underline{k}}}{n^k}k^2\left(\frac{k^{2s-2}}{(2s-2)!!}+O(k^{2s-3})\right)\\&=\frac{n^s}{n^{\underline{s}}}\frac{(2s-1)!!}{(2s-2)!!}Q(n)+O(1)=\frac{(2s-1)!!}{(2s-2)!!}\sqrt{\frac{\pi n}{2}}+O(1).\end{aligned} \tag{2.64}$$

Similarly

$$\begin{aligned}\mathrm{E}(\kappa^2(s))&=\frac{1}{n^{\underline{s}}}\sum_k\frac{n^{\underline{k}}}{n^k}k^3\left\{{k+s-1\atop k}\right\}_s\\&=\frac{1}{n^{\underline{s}}}\sum_k\frac{n^{\underline{k}}}{n^k}k^3\left(\frac{k^{2s-2}}{(2s-2)!!}+O(k^{2s-3})\right)\\&=\frac{n^{s+1}}{n^{\underline{s}}}\frac{(2s)!!}{(2s-2)!!}+O(\sqrt{n})=2sn+O(\sqrt{n}).\end{aligned} \tag{2.65}$$

This implies that, for fixed s, we have

$$\mathrm{var}(\kappa(s))=2sn-\left(\frac{(2s-1)!!}{(2s-2)!!}\right)^2\frac{\pi n}{2}+O(\sqrt{n}). \tag{2.66}$$

For small values of s it is possible to express $\left\{{k+s-1\atop k}\right\}_s$ as a polynomial in k and use equation (7) to compute the exact values of the moments of $\kappa(s)$. The simplest way to find these expansions is to start from the generating function ([Broder84a])

$$\sum_{i\ge0}\left\{{i+r\atop m+r}\right\}_r\frac{z^i}{i!}=\frac{e^{rz}(e^z-1)^m}{m!},\qquad m\ge0.$$

After simple transformations ($i \leftarrow i+m$; $m \leftarrow m-r$) we obtain

$$\sum_{i\geq 0} \left\{ {m+i \atop m} \right\}_r \frac{z^{i+m-r}}{(i+m-r)!} = \frac{e^{rz}(e^z-1)^{m-r}}{(m-r)!}, \qquad m \geq r,$$

so that

$$\left\{ {m+i \atop m} \right\}_r = (m-r+i)^{\underline{i}} \langle z^i \rangle \frac{e^{rz}(e^z-1)^{m-r}}{z^{m-r}}, \qquad m \geq r, \tag{2.67}$$

where the notation $\langle z^i \rangle\, G(z)$ means the coefficient of z^i in the power series expansion of $G(z)$ around $z = 0$.

The last expression can be easily computed within a formal manipulation system by taking the ith derivative with respect to z. Both m and r can be left as symbolic variables.

In our case we need to find $\left\{ {k+s-1 \atop k} \right\}_s$. We obtain

$$\left\{ {k+s-1 \atop k} \right\}_s = (k-1)^{\underline{s-1}} \langle z^{s-1} \rangle \frac{e^{sz}(e^z-1)^{k-s}}{z^{k-s}}, \qquad k \geq s > 0,$$

and in fact this formula holds for any $k \geq 1$ because if $1 \leq k < s$ then $(k-1)^{\underline{s-1}}$ equals 0, and so does $\left\{ {k+s-1 \atop k} \right\}_s$.

The first expansions are

$$\begin{aligned} \left\{ {k+0 \atop k} \right\}_1 &= 1, \\ \left\{ {k+1 \atop k} \right\}_2 &= \frac{1}{2}k^2 + \frac{1}{2}k - 1, \\ \left\{ {k+2 \atop k} \right\}_3 &= \frac{1}{8}k^4 + \frac{5}{12}k^3 - \frac{9}{8}k^2 - \frac{17}{12}k + 2, \end{aligned} \tag{2.68}$$

and from here, via equations (63) and (65),

$$\begin{aligned} \mathbf{E}(\kappa(1)) &= Q(n); \\ \mathbf{var}(\kappa(1)) &= 2n - Q(n)^2 - Q(n); \end{aligned} \tag{2.69}$$

$$\begin{aligned} \mathbf{E}(\kappa(2)) &= \frac{3nQ(n) - n - 2Q(n)}{2(n-1)} = \frac{3}{2}\sqrt{\frac{\pi n}{2}} - 1 + \frac{5}{8}\sqrt{\frac{\pi}{2n}} + O(n^{-1}); \\ \mathbf{var}(\kappa(2)) &= \frac{(32-9\pi)n}{8} - \frac{1}{2}\sqrt{\frac{\pi n}{2}} + \frac{(128-15\pi)}{48} + O(n^{-1/2}); \end{aligned} \tag{2.70}$$

$$\mathrm{E}(\kappa(3)) = \frac{45n^2Q(n) - 25n^2 - 106nQ(n) + 38n + 12Q(n)}{24(n-1)(n-2)}$$
$$= \frac{15}{8}\sqrt{\frac{\pi n}{2}} - \frac{5}{3} + \frac{131}{96}\sqrt{\frac{\pi}{2n}} + O(n^{-1}); \tag{2.71}$$
$$\mathrm{var}(\kappa(3)) = \frac{(768-225\pi)n}{128} - \frac{5}{8}\sqrt{\frac{\pi n}{2}} + \frac{(17408-5895\pi)}{2304} + O(n^{-1}).$$

If both s and n go to infinity such that $s = o(n)$ it can be shown [Pittel83] that

$$\frac{1}{n^{\underline{s}}}\sum_k \frac{n^{\underline{k}}}{n^k}k^2 \left\{ {k+s-1 \atop k} \right\}_s = \sqrt{2sn}(1+o(1)). \tag{2.72}$$

2.6.2. Random starting points

Now let's assume that instead of taking the transitive closure of a fixed set of size s, we start iterating a function f from s starting points chosen uniformly at random, where, as before, f is chosen uniformly at random over $F[n]$. The size of the transitive closure of s points so chosen is now denoted $\kappa(\vec{s})$. This notation is inspired by the fact that the actual number of distinct starting points is only approximately s, and in fact it is a random variable between 1 and s.

Let the chosen points be the sequence $S = (a_1, a_2, \ldots, a_s)$. We encode each function $f \in F[n]$ as a string of length $n+s$ over the alphabet $\{\overline{1}, \ldots, \overline{n}\}$ of the form

$$f \leftrightarrow \overline{f^0(a_1)}\ \overline{f^1(a_1)} \ldots \overline{f^{i_1}(a_1)}$$
$$\overline{f^0(a_2)}\ \overline{f^1(a_2)} \ldots \overline{f^{i_2}(a_2)}$$
$$\ldots \overline{f^0(a_s)}\ \overline{f^1(a_s)} \ldots \overline{f^{i_s}(a_s)}$$
$$\overline{f(b_1)}\ \overline{f(b_2)} \ldots,$$

where i_j is the smallest iterate of f such that $f^i(a_j)$ already appears in the string, and where $b_1, b_2, \ldots$ are the elements of the set $\{1, \ldots, n\} - f^*(S)$, in increasing order.

For example, the encoding of the function in Figure 1, with

$$(a_1, a_2, a_3, a_4) = (3, 1, 3, 6),$$

is

$$f \leftrightarrow \overline{3}\ \overline{10}\ \overline{7}\ \overline{4}\ \overline{9}\ \overline{10}\ \overline{1}\ \overline{9}\ \overline{3}\ \overline{6}\ \overline{1}\ \overline{5}\ \overline{2}\ \overline{1}$$

The resulting encoding can be inverted, reconstructing the function and the starting sequence.

Assume that $|f^*(S)| = k$. That means that, in the encoding, the length of the prefix $\overline{f^0(a_1)}\ldots\overline{f^s(a_s)}$ is exactly $k+s$. The last letter must appear at least twice. Preceding it there is an arbitrary string of length $k+s-1$ over an n letter alphabet, containing k distinct letters. Therefore we can construct a legal prefix as follows:

- Partition the first $k+s-1$ positions into k non-empty subsets. (Each subset corresponds to a certain letter.) This can be done in $\left\{{k+s-1 \atop k}\right\}$ ways.
- Choose k letters and associate each of them to a certain subset. (There are $\binom{n}{k}k!$ possibilities.)
- Choose any letter already chosen for the last position (k possibilities).

Once the prefix is fixed, it can be completed in n^{n-k} ways to form a legal encoding. Hence the probability of reaching k points from s random starting points is

$$\mathbf{Pr}(\kappa(\tilde{s}) = k) = \left\{{k+s-1 \atop k}\right\}\binom{n}{k}k!kn^{-k-s}, \tag{2.73}$$

and the average number of reached points is

$$\mathbf{E}(\kappa(\tilde{s})) = \frac{1}{n^s}\sum_{k\geq 1}\frac{n^{\underline{k}}}{n^k}k^2\left\{{k+s-1 \atop k}\right\}. \tag{2.74}$$

Below are the expected value and the variance of $\kappa(\tilde{s})$ for small s, computed as explained in the previous subsection.

$$\begin{aligned}\mathbf{E}(\kappa(\tilde{1})) &= Q(n);\\ \mathbf{var}(\kappa(\tilde{1})) &= 2n - Q(n)^2 - Q(n);\end{aligned} \tag{2.75}$$

$$\begin{aligned}\mathbf{E}(\kappa(\tilde{2})) &= \frac{3Q(n)-1}{2} = \frac{3}{2}\sqrt{\frac{\pi n}{2}} - 1 + \frac{1}{8}\sqrt{\frac{\pi}{2n}} + O(n^{-1});\\ \mathbf{var}(\kappa(\tilde{2})) &= \frac{16n - 9Q(n)^2 - 8Q(n) + 1}{4}\\ &= \frac{(32-9\pi)n}{8} - \frac{1}{2}\sqrt{\frac{\pi n}{2}} + \frac{32-9\pi}{48} + O(n^{-1/2});\end{aligned} \tag{2.76}$$

$$\begin{aligned}\mathbf{E}(\kappa(\tilde{3})) &= \frac{45nQ(n) - 25n + 2Q(n) + 2}{24n} \\ &= \frac{15}{8}\sqrt{\frac{\pi n}{2}} - \frac{5}{3} + \frac{23}{96}\sqrt{\frac{\pi}{2n}} + O(n^{-1}); \qquad (2.77) \\ \mathbf{var}(\kappa(\tilde{3})) &= \frac{(768 - 225\pi)n}{128} - \frac{5}{8}\sqrt{\frac{\pi n}{2}} + \frac{3584 - 1035\pi}{2304} + O(n^{-1/2}).\end{aligned}$$

2.6.3. Digression – a combinatorial identity

Let S be a set of elements, each chosen uniformly at random. Another way to derive $\mathbf{Pr}(\kappa(\tilde{s}) = k)$ is to start from

$$\begin{aligned}\mathbf{Pr}(\kappa(\tilde{s}) = k) &= \sum_{1 \le i \le s} \mathbf{Pr}(|S| = i)\,\mathbf{Pr}(\kappa(i) = k) \\ &= \sum_i \frac{1}{n^s} \left\{ {s \atop i} \right\} \binom{n}{i} i!\,\mathbf{Pr}(\kappa(i) = k).\end{aligned}$$

Using now equations (61) and (73), we obtain

$$\frac{1}{n^s}\frac{n^{\underline{k}}}{n^k} k \left\{ {k+s-1 \atop k} \right\} = \frac{1}{n^s}\frac{n^{\underline{k}}}{n^k} k \sum_i \left\{ {s \atop i} \right\} \left\{ {k+i-1 \atop k} \right\}_i,$$

and therefore we have the identity

$$\left\{ {k+s-1 \atop k} \right\} = \sum_i \left\{ {s \atop i} \right\} \left\{ {k+i-1 \atop k} \right\}_i.$$

Can we prove it by less intricate methods? The answer is yes and in fact we can prove a more general case.

Theorem 8.

$$\sum_i \left\{ {s \atop i} \right\} \left\{ {m+i \atop k} \right\}_i = \left\{ {m+s \atop k} \right\}.$$

Proof: Consider the $\left\{ {m+s \atop k} \right\}$ partitions of m white balls and s red balls into k non-empty subsets. We can construct them by first partitioning the s red balls into i non-empty subsets, and clumping each subset into a big red ball; then we add the big red balls to the white balls and partition all of them into k non-empty subsets, taking care to keep the big red balls in separate subsets. ∎

2.7. The number of ancestors of one point

We say that y is an *ancestor* of x in f, if there exists an $i \geq 0$ such that $f^i(y) = x$. The number of ancestors of a given point is denoted α. In this section we shall compute the probability distribution of α assuming that all functions in $F[n]$ are equally likely.

Let A be the set of ancestors of a fixed point x. Any mapping such that x has exactly k ancestors can be constructed as follows:

- Choose the $k-1$ elements in $A - \{x\}$. (There are $\binom{n-1}{k-1}$ possibilities.)
- Using all the elements of A, construct a labelled tree rooted at x (k^{k-2} possibilities).
- Choose a random function on the elements not in A. (There are $(n-k)^{n-k}$ possibilities.)
- Choose some value for $f(x)$ (n possibilities)

There are $\binom{n-1}{k-1} n k^{k-2} (n-k)^{n-k}$ ways to carry out this construction, and therefore

$$\Pr(\alpha = k) = \frac{1}{n^n} \binom{n}{k} k^{k-1} (n-k)^{n-k}, \qquad k > 0. \tag{2.78}$$

It is reassuring to notice that by equation (21) we indeed have

$$\sum_{k>0} \frac{1}{n^n} \binom{n}{k} k^{k-1} (n-k)^{n-k} = 1.$$

Remark that the probability that all the points $1, \ldots, n$ are ancestors of a certain element x is just $1/n$, which is the same as the probability that x is a fixed point (see Corollary 1.5). This suggests looking for a bijection between functions where x has n ancestors and functions where x is a fixed point. Here is one possibility. Assume that x has n ancestors in the graph of a certain function f, which means that f has just one cycle and x is included in it. Suppose that the cycle has the form $x \to a_1 \to a_2 \to \cdots \to a_k \to x$. To the function f we associate a function g, identical to f except for the points $a_1, \ldots, a_k$ and x. The cycles of g are built by making x a fixed point, then splitting the string $a_1 a_2 \cdots a_k$ whenever a new maximum is encountered, and then considering each substring as a cycle. This correspondence can be reversed and therefore defines a bijection. In fact, the same idea works for any permutation invariant weight, and therefore we have

Theorem 9. *For any permutation invariant weight the probability that all the elements are ancestors of an element chosen uniformly at random is* $1/n$. ∎

From equations (78) and (23) we obtain that

$$\mathbf{E}(\alpha) = \frac{1}{n^n} \sum_{k>0} \binom{n}{k} k^k (n-k)^{n-k} = Q(n), \tag{2.79}$$

and

$$\begin{aligned} \mathbf{E}(\alpha^2) &= \frac{1}{n^n} \sum_{k\geq 1} \binom{n}{k} k^{k+1} (n-k)^{n-k} = \sum_{k\geq 1} \frac{n^{\underline{k}}}{n^k} \left\{ {k+1 \atop k} \right\} \\ &= \sum_{k\geq 1} \frac{n^{\underline{k}}}{n^k} \frac{k(k+1)}{2} = \frac{nQ(n)+n}{2}. \end{aligned} \tag{2.80}$$

From here

$$\mathbf{var}(\alpha) = \frac{nQ(n) + n - 2Q(n)^2}{2.} \tag{2.81}$$

The expected value for α is not unexpected; it can also be argued as follows. Recall that $\rho(x)$ is the number of descendants of x. For any function f whenever y is a descendant of x, x is an ancestor of y. Hence

$$\sum_x \rho(x,f) = \sum_y \alpha(y,f),$$

and therefore, for any weight distribution, w,

$$\begin{aligned} \mathbf{E}(\rho) &= \sum_f \frac{w(f)}{n} \sum_x \rho(x,f) \\ &= \sum_f \frac{w(f)}{n} \sum_x \alpha(x,f) = \mathbf{E}(\alpha). \end{aligned}$$

In particular for the uniform distribution $\mathbf{E}(\rho) = Q(n)$.

2.8. The number of ancestors of a set of points

Let now A be the set of ancestors of a fixed set of s elements, $S = \{a_1, a_2, \ldots, a_s\}$, with $a_1 < a_2 < \cdots < a_s$. Let $\alpha(s)$ be the size of A. We want to determine the probability distribution of $\alpha(s)$. An arbitrary mapping such that $|A| = k$ can be constructed as follows:

- Choose the $k - s$ elements in $A - S$. (There are $\binom{n-s}{k-s}$ possibilities.)
- Assume for the time being that f is such that $a_1 \to a_2 \to \cdots \to a_s \to a_s$. Choose for the k elements in A, a random mapping having exactly the cycle

given above. It is easy to show via the FF-encoding that there are sk^{k-s-1} such mappings.

- Choose a random mapping for the $n-k$ elements not in A. (There are $(n-k)^{n-k}$ possibilities.)
- Assign arbitrary values to f at the points $a_1, a_2, \ldots, a_s$ (n^s possibilities).

From this construction it follows that

$$\begin{aligned}\Pr(\alpha(s)=k) &= \frac{1}{n^n}\binom{n-s}{k-s} s k^{k-s-1}(n-k)^{n-k} n^s \\ &= \frac{s}{n^{n-s}\underline{n^s}}\binom{n}{k}\underline{k^s} k^{k-s-1}(n-k)^{n-k}.\end{aligned} \tag{2.82}$$

As a quick check, note that

$$\begin{aligned}&\frac{s}{n^{n-s}}\sum_k \binom{n-s}{k-s} k^{k-s-1}(n-k)^{n-k} \\ &\qquad = \frac{s}{n^{n-s}}\sum_k \binom{n-s}{k}(k+s)^{k-1}(n-s-k)^{n-s-k} = 1,\end{aligned}$$

from Abel's identity (equation (20)) when $x \leftarrow s$, $n \leftarrow n-s$, and $y \leftarrow 0$.

For the moments of the distribution of α we compute

$$\begin{aligned}\sum_k \Pr(\alpha(s)=k)k^l &= \frac{s}{n^{n-s}\underline{n^s}}\sum_k \binom{n}{k}\underline{k^s} k^{k-s+l-1}(n-k)^{n-k} \\ &= \frac{s n^s}{\underline{n^s}}\sum_k \frac{\underline{n^k}}{n^k}\left\{{k+l-1 \atop k}\right\}_s,\end{aligned} \tag{2.83}$$

by equation (26). For fixed s and fixed $l > 1$ as $n \to \infty$ this is (equations (1), (8), and (9))

$$\begin{aligned}\sum_k \Pr(\alpha(s)=k)k^l &= \frac{s n^s}{\underline{n^s}}\sum_k \frac{\underline{n^k}}{n^k}\left(\frac{k^{2j-2}}{(2j-2)!!} + O(k^{2j-3})\right) + O(1) \\ &= \frac{s(2j-3)!!\, n^{l-1} Q(n)}{(2j-2)!!} + O(n^{l-2}) \\ &= \frac{s(2j-3)!!\, n^{l-1}}{(2j-2)!!}\sqrt{\frac{\pi n}{2}} + O(n^{l-2}).\end{aligned} \tag{2.84}$$

We can express the first moments as a function of $Q(n)$, by expanding the r-Stirling number as a polynomial in k and adding and subtracting the missing terms. We obtain that

$$\begin{aligned}\mathbf{E}(\alpha(s)) &= \frac{sn^s}{n^{\underline{s}}}\sum_{k\geq s}\frac{n^{\underline{k}}}{n^k} = \frac{sn^s}{n^{\underline{s}}}\left(Q(n) - \sum_{1\leq k<s}\frac{n^k}{n^{\underline{k}}}\right)\\ &= \frac{sn^s}{n^{\underline{s}}}\left(Q(n) - (s-1)(1+O(n^{-1}))\right) = s\sqrt{\frac{\pi n}{2}} + \frac{2s}{3} - s^2 + O(n^{-1/2}),\end{aligned} \tag{2.85}$$

and in a similar manner

$$\mathbf{E}(\alpha(s)^2) = \frac{sn^s}{n^{\underline{s}}}\sum_{k\geq s}\frac{n^{\underline{k}}}{n^k}\left\{{k+1 \atop k}\right\}_s = \frac{sn}{2}\sqrt{\frac{\pi n}{2}} + \frac{sn}{3} + O(\sqrt{n}). \tag{2.86}$$

From the last two equations, with the help of a computer, we obtain

$$\mathbf{var}(\alpha(s)) = \frac{sn}{2}\sqrt{\frac{\pi n}{2}} + \frac{(2s-3\pi s^2)n}{6} + O(\sqrt{n}). \tag{2.87}$$

For small s we can get nicer formulæ:

$$\begin{aligned}\mathbf{E}(\alpha(1)) &= Q(n),\\ \mathbf{var}(\alpha(1)) &= \frac{nQ(n)+n-2Q(n)^2}{2}\\ &= \frac{n}{2}\sqrt{\frac{\pi n}{2}} + \frac{(2-3\pi)n}{6} + \frac{17}{24}\sqrt{\frac{\pi}{2n}} + O(1);\end{aligned} \tag{2.88}$$

$$\begin{aligned}\mathbf{E}(\alpha(2)) &= \frac{2nQ(n)-2n}{n-1}\\ &= 2\sqrt{\frac{\pi n}{2}} - \frac{8}{3} + \frac{13}{6}\sqrt{\frac{\pi}{2n}} + O(n^{-1}),\\ \mathbf{var}(\alpha(1)) &= \frac{n^3Q(n)+n^3-4n^2Q(n)^2+5n^2Q(n)-5n^2+2nQ(n)}{(n-1)^2}\\ &= n\sqrt{\frac{\pi n}{2}} + \frac{(2-6\pi)n}{3} + \frac{39}{4}\sqrt{\frac{\pi}{2n}} + O(1);\end{aligned} \tag{2.89}$$

$$\begin{aligned}\mathbf{E}(\alpha(3)) &= \frac{3n^2Q(n)-6n^2+3n}{(n-1)(n-2)}\\ &= 3\sqrt{\frac{\pi n}{2}} - 7 + \frac{37}{4}\sqrt{\frac{\pi}{2n}} + O(n^{-1}),\\ \mathbf{var}(\alpha(3)) &= \frac{3n}{2}\sqrt{\frac{\pi n}{2}} + \frac{(2-9\pi)n}{2} + \frac{301}{8}\sqrt{\frac{\pi}{2n}} + O(1).\end{aligned} \tag{2.90}$$

Chapter 3

Pollard's factorization method

Let p be a factor of a large integer N. Pollard's Monte Carlo factorization algorithm ([Pollard75],[Brent80]) finds p in average time $O(\sqrt{p})$. Pollard suggested a tuning of his method for the case when a nontrivial factor d of $p-1$ is known, and conjectured that its running time is $O(\sqrt{p/d})$. Pollard and Brent [BP81] used this improved method to factorize the eighth Fermat number, $2^{256}+1$, using $d = 1024$, and recently Gold and Sattler [GS83] ran a series of empirical tests that agree with Pollard's conjecture. Of course, in general, no such d is known, but the improvement is relevant whenever $p-1$ has small factors, whether they are known or not [GS83].

Pollard's method is quite important in practice because although there are several factorization algorithms that are asymptotically faster, they do not take advantage of the existence of small factors. (See [Pomerance82] for a survey; the best current bound is $O(\exp(\sqrt{\ln N \ln\ln N}))$, due to Schnorr and Lenstra [SL84].) Hence it is preferable to use Pollard's method first, to isolate the small factors, and then to switch to a more sophisticated method. Another advantage of this method is that it is extremely simple and can be implemented even on a hand calculator. It is also possible to have a large number of simple processors running Pollard's algorithm on the same composite N, with no communication among processors. (The expected speed-up is discussed in section 6.)

In this chapter, we shall prove Pollard's conjecture under a certain randomness model.

3.1. Pollard's factorization method

A well known method for determining $\lambda(x_0, f)$ and $\mu(x_0, f)$, for a given x_0 is the following algorithm, due to Floyd [Knuth81, ex. 3.1.6]:

$x := x_0$; $y := x_0$; $j := 0$;

repeat { now $x = x_j$ and $y = x_{2j}$ }

$j := j + 1$;

$x := f(x)$;

$y := f(f(y))$;

until $x = y$.

Why does this work? There exists a $j > 0$ such that $x_j = x_{2j}$. The minimum such value is $j = \lceil \mu/\lambda \rceil \lambda$, for $\mu > 0$, and $j = \lambda$ for $\mu = 0$. Hence $j = O(\lambda + \mu)$. Knowing j we can easily determine λ and μ, in time $O(\lambda + \mu)$. Therefore the whole algorithm takes time $O(\lambda + \mu) = O(\rho)$. If all functions are equally likely then according to the results of Chapter 2, $\mathrm{E}(\rho) = \sqrt{\pi n/2} + O(1)$.

There are more efficient algorithms for this problem, based on storing more values of f in memory (see [SSY82] and [Fitch82] for detailed discussions). However, the benefits of the improved versions are not directly applicable to factoring algorithms and do not change the essence of the analysis below.

Pollard's factorization method is based on Floyd's algorithm; f is chosen to be some polynomial $P(x)$ mod N, where N is the number to be factored. The stopping condition is also modified as follows:

$x := x_0$; $y := x_0$; $j := 0$;

repeat { now $x = P^j(x_0) \bmod N$ and $y = P^{2j}(x_0) \bmod N$ }

$j := j + 1$;

$x := P(x) \bmod N$;

$y := P(P(y)) \bmod N$;

until $\gcd(|x - y|, N) > 1$.

Assume that p divides N. By construction $x_{i+1} \equiv P(x_i)$ (modulo N), therefore $x_{i+1} \equiv P(x_i)$ (modulo p). The second congruence implies that for a certain j we have $x_j \equiv x_{2j}$ (modulo p). At this point $\gcd(|x_j - x_{2j}|, N)$ is either N or a proper factor of N. The first case can be shown to be improbable, but if it happens we can try another starting point, or another polynomial P.

So far, we have not discussed what polynomial P to choose. If nothing is known about the factors of N then we can take $P(x) = x^2 + c$, for some constant $c \neq 0$, (see [BP81] and [Knuth81, §4.5.4] for some precautions) but if a factor $d > 2$ of $p - 1$ is known, Pollard [Pollard75] suggests taking $P(x) = x^d + c$. In this case, in the graph of $P(x)$ mod p all the indegrees are either 0 or d (except for c). If no factor of $p-1$ is known, but $p-1$ contains small factors, we can use $P(x) = x^a + c$, with a a product of small primes. Assume that $\gcd(p-1, a) = d > 1$; then again, in the graph of $P(x)$ mod p all the indegrees are either 0 or d (except for c). Pollard conjectured that in this case $\mathbf{E}(\rho) = O\left(\sqrt{n/(d-1)}\right)$, a $\sqrt{d-1}$ improvement over an arbitrary choice for $P(x)$, at a cost of $O(\log d)$ more operations per iteration.

To determine the expected running time of the algorithm we must compute the expected cycle length, and the expected tail length in such a mapping. As in other analyses of factorization algorithms, we assume that all such mappings (that is, where all the indegrees are either 0 or d) are equiprobable. Under this model, we shall prove that Pollard's conjecture is true. Deciding on the validity of such a model is beyond the current state of knowledge in number theory, but experimental results ([BP81], [Knuth81], [GS83], [Pollard83]) seem to confirm it.

Variants of this algorithm (e.g., [Brent80]) depend in a slightly different way on λ and μ, but their running time is still essentially proportional to ρ.

3.2. The constant indegree model

Let's consider the family $\mathcal{F}$ of functions $f : \{1, \ldots, n\} \mapsto \{1, \ldots, n\}$ such that exactly n/d nodes have indegree d. (Here n represents the number $p - 1$ in the factorization problem.) This family is not empty only if n/d is an integer, say m; then $\mathcal{F}$ has cardinality $\binom{n}{m} n!/(d!)^m$. We define a permutation invariant probability weight as follows: to each function in $\mathcal{F}$ we assign probability $(d!)^m/(\binom{n}{m} n!)$ and to all other functions $f : \{1, \ldots, n\} \mapsto \{1, \ldots, n\}$ we assign probability 0. (In fact this weight is strongly invariant.)

For this probability weight it is easy to see that

$$\mathbf{Pr}(\nu \geq k) = \frac{1}{|\mathcal{F}|} \binom{n}{m} \binom{m}{k} k! \frac{(n-k)!}{((d-1)!)^k (d!)^{m-k}}, \tag{3.1}$$

because the trivial encoding of a function f with $\nu \geq k$ can be constructed as follows:

- Choose the m elements with indegree d. (There are $\binom{n}{m}$ possibilities.)
- Choose k elements out of these m elements, to be the first k letters of $C_0(f)$. (There are $\binom{m}{k}$ ways.)

- Permute the first k elements (in one of $k!$ ways).
- The remaining $n-k$ letters are a permutation of k letters repeated $d-1$ times and $m-k$ letters repeated d times. (Hence the number of possibilities is $(n-k)!\,((d-1)!)^{-k}(d!)^{-(m-k)}$.)

Expanding equation (1) we obtain

$$\begin{aligned}\mathbf{Pr}(\nu \geq k) &= \binom{m}{k}\frac{k!\,(n-k)!\,(d!)^m}{((d-1)!)^k (d!)^{m-k} n!} \\ &= \frac{\binom{m}{k}d^k}{\binom{n}{k}} = \frac{(n-k)!\,m!\,d^k}{n!\,(m-k)!} = \frac{\binom{n-k}{m-k}d^k}{\binom{n}{m}}.\end{aligned} \tag{3.2}$$

We are interested in the moments of the distribution of r. To obtain them we shall consider the generating function

$$F(z) = \sum_{k\geq 0} \mathbf{Pr}(r \geq k) z^k. \tag{3.3}$$

Clearly, the probability generating function of r, $C(z)$, satisfies

$$C(z) = F(z) - \frac{1}{z}(F(z) - 1); \tag{3.4}$$

hence

$$\begin{aligned} C'(1) &= F(1) - 1, \\ C''(1) &= 2F'(1) - 2F(1) + 2, \\ C'''(1) &= 3F''(1) - 6F'(1) + 6F(1) - 6, \end{aligned} \tag{3.5}$$

and so on. Now to obtain the derivatives of F we first write

$$F(z) = G(zd)/\binom{n}{m}, \tag{3.6}$$

where

$$G(z) = \sum_{k\geq 0} \binom{n-k}{m-k} z^k = \sum_{k\leq m} \binom{n-m+k}{k} z^{m-k} \tag{3.7}$$

This is related to the tail of a negative binomial distribution that can be replaced by the tails of a binomial distribution, via the following theorem.

Theorem 1. *Let p be a probability, and let $q = 1 - p$. We have*

$$\sum_{k\le m}\binom{r+k-1}{k}p^r q^k = \sum_{k\le m}\binom{r+m}{k}p^{r+m-k}q^k.$$

Proof: We may assume that r is an integer. The term $\binom{r+k-1}{k}p^r q^k$ is the probability to obtain the rth success in a sequence of Bernoulli trials, after exactly $r+k$ trials. So that the left side is the probability to obtain the rth success in at most $r+m$ trials = the probability to fail at most m times in $r+m$ trials.[1] ∎

Corollary 2. *If m is an integer and x, y, r are arbitrary,*

$$\sum_{k\le m}\binom{r+k-1}{k}x^k(x+y)^{m-k} = \sum_{k\le m}\binom{r+m}{k}x^k y^{m-k}.$$

Proof: Multiply both sides of the identity in Theorem 1 by $(x+y)^m/p^r$, and then replace $p \leftarrow y/(x+y)$ and $q \leftarrow x/(x+y)$. ∎

Applying Corollary 2 to equation (7), with $r \leftarrow n-m+1$, $x \leftarrow 1$, $y \leftarrow z-1$ we get

$$G(z) = \sum_{k\le m}\binom{n+1}{k}(z-1)^{m-k}, \tag{3.8}$$

and hence

$$G^{(l)}(z) = \sum_{k\le m}\binom{n+1}{k}(m-k)^{\underline{l}}(z-1)^{m-k-l}. \tag{3.9}$$

Now we use the expansion[2]

$$(m-k)^{\underline{l}} = \sum_i \binom{l}{i}(-1)^i(m-i)^{\underline{l-i}}k^{\underline{i}}, \tag{3.10}$$

---

[1] The standard proof [Pearson33] of this theorem is to use calculus of complex variables to equate both sides to the tails of the Beta distribution. However after finding this proof, a careful search of the literature showed that it was already published twenty five years ago [Patil60].

[2] *Proof:* We repeatedly use the identities $x^{\overline{n}} = (x+n-1)^{\underline{n}}$ and $(-x)^{\underline{n}} = (-1)^n x^{\overline{n}}$ to obtain

$$\begin{aligned}(m-k)^{\underline{l}} &= (-1)^l(k-m)^{\overline{l}} = (-1)^l(k-m+l-1)^{\underline{l}}\\ &= (-1)^l\sum_i \binom{l}{i}k^{\underline{i}}(-m+l-1)^{\underline{l-i}}\\ &= \sum_i \binom{l}{i}k^{\underline{i}}(-1)^i(m-l+1)^{\overline{l-i}} = \sum_i \binom{l}{i}k^{\underline{i}}(-1)^i(m-i)^{\underline{l-i}}.\end{aligned}$$

to obtain

$$G^{(l)}(z) = \sum_i \binom{l}{i}(-1)^i(m-i)^{\underline{l-i}}(n+1)^{\underline{i}} \sum_{k\le m} \binom{n+1-i}{k-i}(z-1)^{m-k-l}$$
$$= \sum_i \binom{l}{i}(-1)^i(m-i)^{\underline{l-i}}(n+1)^{\underline{i}} \sum_{k\le m-i} \binom{n+1-i}{k}(z-1)^{m-i-k-l}. \tag{3.11}$$

Applying Corollary 2 again, this time in reverse, yields

$$G^{(l)}(z) = \sum_i \binom{l}{i}(-1)^i(m-i)^{\underline{l-i}}(n+1)^{\underline{i}} \sum_{k\le m-i} \binom{n-m+k}{k} z^{m-i-k}(z-1)^{-l}$$
$$= \frac{1}{(z-1)^l} \sum_i \binom{l}{i}(-1)^i(m-i)^{\underline{l-i}}(n+1)^{\underline{i}} z^{-i} S_i(z), \tag{3.12}$$

where

$$S_i(z) = \sum_{k\le m-i} \binom{n-m+k}{k} z^{m-k} = S_0(z) - \sum_{0\le k<i} \binom{n-k}{m-k} z^k. \tag{3.13}$$

We are most interested in $G^{(l)}(z)$ when l is small, hence the formulæ that we have derived are actually "simple" to compute, in spite of their forbidding appearance. We have

$$G(z) = S_0(z);$$
$$G'(z) = \frac{mS_0(z) - (n+1)z^{-1}S_1(z)}{z-1}$$
$$= \frac{(mz-n-1)S_0(z) + (n+1)\binom{n}{m}}{z(z-1)};$$
$$G''(z) = \frac{m(m-1)S_0(z) - 2(m-1)(n+1)z^{-1}S_1(z) + n(n+1)z^{-2}S_2(z)}{(z-1)^2}. \tag{3.14}$$

Let $A = S_0(d)/\binom{n}{m}$. Since $F^{(l)}(1) = d^l G^{(l)}(d)/\binom{n}{m}$ and $n = md$, we find that these equations simplify considerably:

$$F(1) = A;$$
$$F'(1) = \frac{n+1-A}{d-1}; \tag{3.15}$$
$$F''(1) = \frac{((d-1)n+2d)A - 2(n+1)d}{(d-1)^2}.$$

It remains to determine the asymptotic value of $A = G(d)/\binom{n}{m}$. By equation (8) we can write this as

$$\begin{aligned} A &= \frac{1}{\binom{n}{m}} \sum_{k \le m} \binom{n+1}{k} (d-1)^{m-k} \\ &= \frac{1}{\binom{n}{m}} \frac{d^{n+1}}{(d-1)^{n-m+1}} \sum_{k<(n+1)/d} \binom{n+1}{k} \left(\frac{1}{d}\right)^k \left(1-\frac{1}{d}\right)^{n+1-k}. \end{aligned} \tag{3.16}$$

The sum is $1/2 + O(n^{-1/2})$, by the central limit theorem, since it is the sum of all probabilities that are less than the mean value $(n+1)/d$ of a binomial distribution. And the leading coefficient is easy to evaluate by Stirling's approximation:

$$\begin{aligned} \frac{1}{\binom{md}{m}} \frac{d^{md+1}}{(d-1)^{md-m+1}} &= \frac{\sqrt{2\pi m}\, m^m \sqrt{2\pi m(d-1)}\, m^{m(d-1)}}{\sqrt{2\pi md}\, m^{md}} \frac{d}{d-1}\left(1+O(m^{-1})\right) \\ &= \sqrt{\frac{2\pi md}{d-1}}\left(1+O(m^{-1})\right). \end{aligned} \tag{3.17}$$

Hence

$$A = \sqrt{\frac{\pi n}{2(d-1)}} + O(1). \tag{3.18}$$

This analysis is sufficient to give the leading term in our asymptotic formulæ, but it is somewhat unsatisfactory since it does not make clear how we could obtain better accuracy. For example, we might want to know the constant term of A. The next section sharpens the asymptotics by looking closer at the left half of a binomial distribution.

3.3. Sums of Bernoulli random variables

We start from the following theorem due to Esseen [Esseen45]. (See also [GK68] and [Petrov75].)

Theorem 3. *If $\xi_1, \xi_2, \ldots, \xi_n$ are independent, identically distributed, random variables, with mean 0, variance σ^2, and finite third moment, α_3, such that the only possible values of ξ_k are $a + \nu h$ for $\nu = 0, \pm 1, \pm 2, \ldots$, and h is maximum, then the cumulative probability distribution*

$$F_n(x) = \mathbf{Pr}\left(\textstyle\sum_i \xi_i/(\sqrt{n}\,\sigma) < x\right)$$

satisfies

$$F_n(x) = \Phi(x) + \frac{e^{-x^2/2}}{\sqrt{2\pi n}}\left(\frac{(1-x^2)\alpha_3}{6\sigma^3} + \frac{h}{\sigma}S\left(\frac{x\sigma\sqrt{n}}{h} - \frac{na}{h} + \left\lceil\frac{na}{h}\right\rceil\right)\right) + o(n^{-1/2}),$$

where $\Phi(x)$ *is the normal distribution,*

$$\Phi(x) = \frac{1}{2\pi}\int_{-\infty}^{x} e^{-t^2/2}\,dt,$$

and $S(x)$ *is a discontinuous function*[3] *of* x,

$$S(x) = \lceil x\rceil - x - 1/2.$$

∎

We are interested in the sum of n Bernoulli variables $\chi_1, \chi_2, \ldots, \chi_n$, with $\mathbf{Pr}(\chi_k = 1) = p$. Let $\xi_k = \chi_k - p$, and $q = 1 - p$. Then $\mathbf{E}(\xi_k) = 0$, $\mathbf{var}(\xi_k) = pq$, and $\alpha_3(\xi_k) = pq^3 - qp^3 = pq(q-p)$. We can apply Esseen's theorem with $a = -p$ and $h = 1$ to obtain

$$\begin{aligned}&\mathbf{Pr}\left(\frac{\sum_k \xi_k}{\sqrt{npq}} < x\right)\\ &\quad = \Phi(x) + \frac{e^{-x^2/2}}{\sqrt{2\pi n}}\left(\frac{(1-x^2)(q-p)}{6\sqrt{pq}} + \frac{S(x\sqrt{npq}+np)}{\sqrt{pq}}\right) + O(n^{-1}).\end{aligned} \tag{3.19}$$

(We are allowed in this case to replace $o(n^{-1/2})$ by $O(n^{-1})$ because the fourth moment is also finite. For details see [Esseen45] or [Petrov75].)

Making $x = 0$ and substituting $\chi_k - p$ for ξ_k we obtain that

$$\begin{aligned}\sum_{k<np}\binom{n}{k}p^k q^{n-k} &= \Phi(0) + \frac{1}{\sqrt{2\pi npq}}\left(\frac{q-p}{6} + \lceil np\rceil - np - \frac{1}{2}\right) + O(n^{-1})\\ &= \frac{1}{2} + \frac{1}{\sqrt{2\pi npq}}\left(\lceil np\rceil - np - \frac{p+1}{3}\right) + O(n^{-1}).\end{aligned} \tag{3.20}$$

---

[3] The literature is a bit confusing with respect to this function, so maybe some clarification is necessary. In the original paper by Esseen, $F_n(x)$ is probably meant to be $\mathbf{Pr}(\sum_i \xi_i/(\sqrt{n}\sigma) \le x)$. There is no definition of $F_n(x)$ but there is a picture that implies that $F_n(x)$ is continuous from the *right*. Esseen uses instead of S, the function $\bar{S} = \lfloor x\rfloor - x + 1/2$. Notice that $S(x) = \bar{S}(x)$ for all non-integral x; at integral points $S(x)$ is continuous from the left, while $\bar{S}(x)$ is continuous from the right. In [GK68] the cumulative probability distribution, $F_n(x)$, is defined as above, but the authors incorrectly use $\bar{S}$. In [Petrov75] the function S is described by its Fourier expansion only. Of course S and $\bar{S}$ have the same Fourier expansion ... and in fact this is the source of error in [GK68].

3.4. Better asymptotics

Setting $n \leftarrow n+1$, $p \leftarrow 1/d$, and $q \leftarrow 1-1/d$, in equation (20) we can improve the estimate of the sum in equation (16):

$$\begin{aligned}\sum_{k<(n+1)/d} \binom{n+1}{k} \left(\frac{1}{d}\right)^k \left(1-\frac{1}{d}\right)^{n+1-k} \\ = \frac{1}{2} + \frac{d}{\sqrt{2\pi n(d-1)}} \left(\left\lceil \frac{n+1}{d} \right\rceil - \frac{n+1}{d} - \frac{1/d+1}{3} \right) + O(n^{-1}) \\ = \frac{1}{2} + \frac{1}{\sqrt{2\pi n(d-1)}} \left(\frac{2d}{3} - \frac{4}{3} \right) + O(n^{-1}).\end{aligned} \tag{3.21}$$

Hence we obtain a more precise value for A:

$$A = \sqrt{\frac{\pi n}{2(d-1)}} + \frac{2(d-2)}{3(d-1)} + O(n^{-1/2}). \tag{3.22}$$

From here, using the equations (15), we obtain

$$\begin{aligned} F(1) &= \sqrt{\frac{\pi n}{2(d-1)}} + \frac{2(d-2)}{3(d-1)} + O(n^{-1/2}); \\ F'(1) &= \frac{n}{d-1} - \sqrt{\frac{\pi n}{2(d-1)^3}} + \frac{d+1}{3(d-1)^2} + O(n^{-1/2}); \\ F''(1) &= \sqrt{\frac{\pi n^3}{2(d-1)^3}} - \frac{4(d+1)n}{3(d-1)^2} + O(n^{1/2}). \end{aligned} \tag{3.23}$$

Finally, going back to the equations (5) and to the relevant equations in Chapter 1, we obtain

$$\begin{aligned} \mathbf{E}(r) &= \sqrt{\frac{\pi n}{2(d-1)}} - \frac{d+1}{3(d-1)} + O(n^{-1/2}), \\ \mathbf{var}(r) &= \frac{(4-\pi)n}{2(d-1)} - \frac{d+1}{3}\sqrt{\frac{\pi n}{2(d-1)^3}} + O(1); \end{aligned} \tag{3.24}$$

$$\begin{aligned} \mathbf{E}(\lambda) &= \frac{1}{2}\sqrt{\frac{\pi n}{2(d-1)}} + \frac{d-2}{3(d-1)} + O(n^{-1/2}), \\ \mathbf{var}(\lambda) &= \frac{(16-3\pi)n}{24(d-1)} - \frac{d+1}{6}\sqrt{\frac{\pi n}{2(d-1)^3}} + O(1); \end{aligned} \tag{3.25}$$

$$\begin{aligned} \mathbf{E}(\mu) &= \frac{1}{2}\sqrt{\frac{\pi n}{2(d-1)}} + \frac{d-5}{3(d-1)} + O(n^{-1/2}), \\ \mathbf{var}(\mu) &= \frac{(16-3\pi)n}{24(d-1)} - \frac{d+1}{6}\sqrt{\frac{\pi n}{2(d-1)^3}} + O(1); \end{aligned} \tag{3.26}$$

$$\begin{aligned} \mathbf{E}(\rho) &= \sqrt{\frac{\pi n}{2(d-1)}} + \frac{2d-7}{3(d-1)} + O(n^{-1/2}), \\ \mathbf{var}(\rho) &= \frac{(4-\pi)n}{24(d-1)} - \frac{d+1}{3}\sqrt{\frac{\pi n}{2(d-1)^3}} + O(1); \end{aligned} \tag{3.27}$$

$$\mathbf{cov}(\lambda,\mu) = \frac{(8-3\pi)n}{24(d-1)} + O(1). \tag{3.28}$$

These values confirm the constant term in a sharper form of Pollard's conjecture [Pollard82]. In principle, smaller order terms and higher order moments can be computed by the same method, using smaller order terms in the estimate of the tails of the binomial distribution.

3.5. The case $d = 2$

This case is of special interest for two reasons: it corresponds to the frequently used polynomial $x^2 + c$ in Pollard's method and we can obtain closed form formulæ that are a useful check on the general case.

If $d = 2$, by equation (13) and Corollary 2, with $r \leftarrow m+1$, $x \leftarrow 1$, and $y \leftarrow 1$, the sums S_i are given by

$$\begin{aligned} S_i(2) &= \sum_{k\le m}\binom{m+k}{k}2^{m-k} - \sum_{0\le k<i}\binom{2m-k}{m-k}2^k \\ &= \sum_{k\le m}\binom{2m+1}{k} - \sum_{0\le k<i}\binom{2m-k}{m-k}2^k \\ &= 2^{2m} - \sum_{0\le k<i}\binom{2m-k}{m-k}2^k. \end{aligned} \tag{3.29}$$

From here

$$F(1) = \frac{2^n}{\binom{n}{n/2}} = \sqrt{\frac{\pi n}{2}} + \frac{1}{4}\sqrt{\frac{\pi}{2n}} + O(n^{-3/2}),$$

$$F'(1) = n - \frac{2^n}{\binom{n}{n/2}} + 1 = n - \sqrt{\frac{\pi n}{2}} + 1 + O(n^{-1}), \tag{3.30}$$

$$F''(1) = \frac{2^n n}{\binom{n}{n/2}} - 4n + \frac{2^{n+2}}{\binom{n}{n/2}} - 4 = \sqrt{\frac{\pi n^3}{2}} - 4n + O(n^{1/2}).$$

(We have used the expansion

$$\frac{2^{2m}}{\binom{2m}{m}} = \sqrt{\pi m}\left(1 + \frac{1}{8m} + \frac{1}{128m^2} + O(m^{-3})\right), \tag{3.31}$$

which is easy to compute using Stirling's formula.)

Going back one more step, we finally obtain

$$\begin{aligned}
\mathbf{E}(r) &= \frac{2^n}{\binom{n}{n/2}} - 1 = \sqrt{\frac{\pi n}{2}} - 1 + \frac{1}{4}\sqrt{\frac{\pi}{2n}} + O(n^{-3/2}), \\
\mathbf{var}(r) &= 2n - \frac{2^{2n}}{\binom{n}{n/2}^2} - \frac{2^n}{\binom{n}{n/2}} + 2 \\
&= \frac{(4-\pi)n}{2} - \sqrt{\frac{\pi n}{2}} + \frac{(8-\pi)}{4} + O(n^{-1});
\end{aligned} \tag{3.32}$$

$$\begin{aligned}
\hat{\mathbf{E}}(\lambda) &= \frac{2^{n-1}}{\binom{n}{n/2}} = \frac{1}{2}\sqrt{\frac{\pi n}{2}} + \frac{1}{8}\sqrt{\frac{\pi}{2n}} + O(n^{-3/2}), \\
\widehat{\mathbf{var}}(\lambda) &= \frac{2n}{3} - \frac{2^{2n-2}}{\binom{n}{n/2}^2} - \frac{2^{n-1}}{\binom{n}{n/2}} + \frac{2}{3} \\
&= \frac{(16-3\pi)n}{24} - \frac{1}{2}\sqrt{\frac{\pi n}{2}} + \frac{(32-3\pi)}{48} + O(n^{-1});
\end{aligned} \tag{3.33}$$

$$\begin{aligned}
\hat{\mathbf{E}}(\mu) &= \frac{2^{n-1}}{\binom{n}{n/2}} - 1 + \frac{2^n}{n\binom{n}{n/2}} - \frac{1}{n} \\
&= \frac{1}{2}\sqrt{\frac{\pi n}{2}} - 1 + \frac{9}{8}\sqrt{\frac{\pi}{2n}} - \frac{1}{n} + O(n^{-3/2}), \\
\widehat{\mathbf{var}}(\mu) &= \frac{(16-3\pi)n}{24} - \frac{1}{2}\sqrt{\frac{\pi n}{2}} + \frac{(128-27\pi)}{48} + O(n^{-1});
\end{aligned} \tag{3.34}$$

$$\begin{aligned}\hat{\mathrm{E}}(\rho) &= \frac{2^n}{\binom{n}{n/2}} - 1 + \frac{2^n}{n\binom{n}{n/2}} - \frac{1}{n} \\ &= \sqrt{\frac{\pi n}{2}} - 1 + \frac{5}{4}\sqrt{\frac{\pi}{2n}} - \frac{1}{n} + O(n^{-3/2}), \qquad (3.35) \\ \widehat{\mathrm{var}}(\rho) &= \frac{(4-\pi)n}{2} - \sqrt{\frac{\pi n}{2}} + \frac{(24-5\pi)}{4} + O(n^{-1});\end{aligned}$$

$$\begin{aligned}\widehat{\mathrm{cov}}(\lambda,\mu) &= \frac{n}{3} - \frac{2^{2n-2}}{\binom{n}{n/2}^2} + \frac{4}{3} - \frac{2^{2n-1}}{n\binom{n}{n/2}^2} - \frac{2^{n-1}}{n\binom{n}{n/2}} + \frac{1}{n} \\ &= \frac{(8-3\pi)n}{24} + \frac{64-15\pi}{48} + O(n^{-1}).\end{aligned} \qquad (3.36)$$

Comparing these with the corresponding formulæ for the uniform case (equations (2.35) and following) shows that the leading terms are unaffected, but the next terms decrease very slightly.

3.6. The parallelization of Pollard's factorization algorithm

Suppose that we have s processors simultaneously running Pollard's algorithm trying to factor the same number. What is the expected speed-up? More precisely, what is the expected value of the minimum running time to completion, over the s processors? As a model, let's assume that processor i computes the length of the period and of the cycle of a random function $f_i \in F[n]$ (in fact a polynomial) starting from the point x_i. We need to compute

$$\mathrm{E}\big(\min\big(\rho(f_1,x_1),\rho(f_2,x_2),\ldots,\rho(f_s,x_s)\big)\big).$$

We consider two cases. The first case is that each processor chooses its function uniformly at random over the n^n possible functions. Then the following theorem applies.

Theorem 4. *Let $f_1, f_2, \ldots, f_s : \{1,\ldots,n\} \mapsto \{1,\ldots,n\}$ be s mappings chosen uniformly at random in $F[n]$. For any fixed choice of $x_1, x_2, \ldots, x_s$*

$$\mathrm{E}\big(\min\big(\rho(f_1,x_1),\rho(f_2,x_2),\ldots,\rho(f_s,x_s)\big)\big) = \sum_{k\geq 1}\left(\binom{n}{k}\frac{k!}{n^k}\right)^s = \sqrt{\frac{\pi n}{2s}} + O(1).$$

Proof: From the results of Section 2.4 it follows that if f_i is chosen uniformly at random, then

$$\mathbf{Pr}(\rho(f_i, x_i) \geq k) = \mathbf{Pr}(\nu(f_i) \geq k) = \frac{n^{\underline{k}}}{n^k}.$$

Because $\rho(f_1, x_1), \rho(f_2, x_2), \ldots, \rho(f_s, x_s)$ are independent random variables, we have

$$\mathbf{Pr}(\min(\rho(f_1, x_1), \rho(f_2, x_2), \ldots, \rho(f_s, x_s)) \geq k) = \left(\frac{n^{\underline{k}}}{n^k}\right)^s,$$

and

$$\mathbf{E}(\min(\rho(f_1, x_1), \rho(f_2, \tau_2), \ldots, \rho(f_s, x_s))) = \sum_{k \geq 1} \left(\frac{n^{\underline{k}}}{n^k}\right)^s.$$

For a fixed s and every k,

$$\left(\frac{n^{\underline{k}}}{n^k}\right)^s = \prod_{0 \leq j < k} \left(1 - \frac{j}{n}\right)^s = \prod_{0 \leq j < k} \left(1 - \frac{js}{n} + O\left(\frac{j^2}{n^2}\right)\right)$$
$$= \prod_{0 \leq j < k} \left(1 - \frac{sj}{n}\right)\left(1 + O\left(\frac{j^2}{n^2}\right)\right) = \frac{(n/s)^{\underline{k}}}{(n/s)^k}\left(1 + O\left(\frac{k^3}{n^2}\right)\right)$$

If $k \geq n/s$ then the term $n^{\underline{k}}/n^k$ is clearly exponentially small and therefore

$$\sum_{k \geq 1} \left(\frac{n^{\underline{k}}}{n^k}\right)^s = \sum_{1 \leq k < n/s} \left(\frac{n^{\underline{k}}}{n^k}\right)^s + \text{exponentially small terms}$$
$$= \sum_{1 \leq k < n/s} \frac{(n/s)^{\underline{k}}}{(n/s)^k}\left(1 + O\left(\frac{k^3}{n^2}\right)\right)$$
$$= \sum_{1 \leq k < n/s} \frac{(n/s)^{\underline{k}}}{(n/s)^k} + O\left(\sum_{1 \leq k < n/s} \frac{(n/s)^{\underline{k}}}{(n/s)^k}\frac{k^3}{n^2}\right)$$

It is easy to see that

$$Q(\lfloor n/s \rfloor) \leq \sum_{1 \leq k < n/s} \frac{(n/s)^{\underline{k}}}{(n/s)^k} \leq Q(\lceil n/s \rceil),$$

and hence

$$\sum_{1 \leq k < n/s} \frac{(n/s)^{\underline{k}}}{(n/s)^k} = \sqrt{\frac{\pi n}{2s}} + O(1).$$

In a similar manner

$$\frac{1}{n^2}Q_{\lfloor n/s\rfloor}(1,2^3,3^3,\ldots) \le \sum_{1\le k<n/s}\frac{(n/s)^{\underline{k}}}{(n/s)^k}\frac{k^3}{n^2} \le \frac{1}{n^2}Q_{\lceil n/s\rceil}(1,2^3,3^3,\ldots),$$

and from equation (2.8) it follows that

$$\sum_{1\le k<n/s}\frac{(n/s)^{\underline{k}}}{(n/s)^k}\frac{k^3}{n^2} = \frac{2}{s^2}+O(n^{-1/2}).$$

Putting everything back together, we obtain that

$$\sum_{k\ge 1}\left(\frac{n^{\underline{k}}}{n^k}\right)^s = \sqrt{\frac{\pi n}{2s}}+O(1).$$

∎

Hence if every processor uses an independently chosen random function then the speed-up is $O(\sqrt{s})$, where s is the number of processors. (Compare with equation (2.41).) The second case to consider is that all processors use the same function but different starting points. However, via very general principles, it can be shown that this strategy is never better than the first strategy.

We need now to compute

$$\mathbf{E}\big(\min(\rho(f,x_1),\rho(f,x_2),\ldots,\rho(f,x_s))\big),$$

where f is chosen uniformly at random in $F[n]$ and $x_1,x_2,\ldots,x_s$ are chosen uniformly at random in $\{1,\ldots,n\}$.

Define $b(f,k)$ to be the probability that $\rho(f,x)\ge k$ when x is chosen uniformly at random. This means that $nb(f,k)$ is the number of (bad!) points x such that $\rho(f,x)\ge k$. Then the probability that the second strategy requires more than k steps is

$$\mathbf{Pr}\big(\min(\rho(f,x_1),\rho(f,x_2),\ldots,\rho(f,x_s))\ge k\big) = \sum_{f\in F[n]}\frac{b(f,k)^s}{n^n}, \qquad (3.37)$$

while the probability (which we already computed) that the first strategy requires more than k steps is

$$\Pr\bigl(\min(\rho(f_1,x_1),\rho(f_2,x_2),\ldots,\rho(f_s,x_s)) \geq k\bigr) = \Bigl(\sum_{f\in F[n]} \frac{b(f,k)}{n^n}\Bigr)^s. \qquad (3.38)$$

We shall now show that

$$\sum_{f\in F[n]} \frac{\bigl(b(f,k)\bigr)^s}{n^n} \geq \Bigl(\sum_{f\in F[n]} \frac{b(f,k)}{n^n}\Bigr)^s, \qquad (3.39)$$

regardless of the actual values of $b(f,k)$, and therefore

$$\begin{aligned}&\mathbf{E}\bigl(\min(\rho(f,x_1),\rho(f,x_2),\ldots,\rho(f,x_s))\bigr)\\ &\qquad\geq \mathbf{E}\bigl(\min(\rho(f_1,x_1),\rho(f_2,x_2),\ldots,\rho(f_s,x_s))\bigr).\end{aligned} \qquad (3.40)$$

We start from

Theorem 5. *Let $x_1 \leq x_2 \leq \cdots \leq x_m$ be m real points. Any real function f such that $f''(x)$ exists and $f''(x) > 0$ on the interval $[x_1, x_m]$ satisfies*

$$\sum_{1\leq i\leq m} \frac{f(x_i)}{m} \geq f\Bigl(\sum_{1\leq i\leq m} \frac{x_i}{m}\Bigr).$$

Proof: See [HLP59] page 72. ∎

Applying now Theorem 5 to the function $f(x) = x^s$, we obtain that for any collection of m positive points, $x_1, x_2, \ldots, x_m$, we have

$$\sum_{1\leq i\leq m} \frac{x_i^s}{m} \geq \Bigl(\sum_{1\leq i\leq m} \frac{x_i}{m}\Bigr)^s, \qquad (3.41)$$

and in particular

$$\sum_{f\in F[n]} \frac{\bigl(b(f,k)\bigr)^s}{n^n} \geq \Bigl(\sum_{f\in F[n]} \frac{b(f,k)}{n^n}\Bigr)^s, \qquad (3.42)$$

which is the inequality we wanted to prove.

In conclusion, if s processors are running Pollard's algorithm in parallel, they should run it with different polynomials, for an expected speed-up of $\sqrt{s}$. The strategy of using the same polynomial and different starting points is inferior on average. Although the gain is relatively small compared with the number of processors used, the parallel version of Pollard's algorithm might be a good choice in certain situations (e.g., vector machines) because no communication is required.

3.7. Open problems

The obvious, but probably hopeless question is how accurately the model used here for Pollard's method reflects reality.

Another problem, more amenable to solution, is to compute the expected values of λ and μ if every node has indegree either a, or b, or 0; more generally one can consider a given indegree probability distribution, or other distributions closely related to the polynomials that are actually used.

References

[Abel1826] N. H. Abel, "Beweis eines Ausdruckes, von welchem die Binomial-Formel ein einzelner Fall ist," *Journal für die reine und angewandte Mathematik*, **1**(1826), 159–160.

[Berg81] S. Berg, "On snowball sampling, random mappings and related problems," *Journal of Applied Probability*, **18**(1981), 283–290.

[BP81] R. P. Brent and J. M. Pollard, "Factorization of the Eighth Fermat Number," *Mathematics of Computation*, **36**(1981), 627–630.

[Brent80] R. P. Brent, "An improved Monte Carlo Factorization Algorithm," *BIT*, **20**(1980), 176–184.

[Broder83] A. Z. Broder, "A general expression for Abelian identities," in: L. J. Cummings (ed.), *Combinatorics on Words*, Academic Press, 1983.

[Broder84a] A. Z. Broder, "The r-Stirling numbers," *Discrete Mathematics*, **49**(1984), 241–259.

[Broder84b] A. Z. Broder, "Two counting problems solved via string encodings," in: A. Apostolico and Z. Galil, *Combinatorial Algorithms on Words*, Springer-Verlag, to appear.

[Carlitz80] L. Carlitz, "Weighted Stirling numbers of the first and second kind," *The Fibonacci Quarterly*, **18**(1980), 147–162, 242–257.

[Cauchy1826] A. Cauchy, *Exercises de Mathématiques*, Paris, 1826.

[Comtet74] L. Comtet, *Advanced Combinatorics*, Reidel, Dordrecht/Boston, 1974.

[Esseen45] C. G. Esseen, "Fourier analysis of distribution functions," *Acta Mathematica*, **77**(1945), 1–125.

[FF70] D. Foata and A. Fuchs, "Réarrangements des fonctions et dénombrement," *Journal of Combinatorial Theory*, **8**(1970), 361–375.

[Fitch72] F. E. Fitch, "Two problems in concrete complexity: Cycle detection and parallel prefix computation," Research Report RJ3651, IBM Research Laboratory, San Jose, Calif., 1982.

[Foata65] D. Foata, "Etude algébrique de certain problèmes d'analyse combinatoire et du calcul des probabilités," *Publications de l'Institut de Statistique de l'Université de Paris*, **14**(1965), 81–241.

[Foata83] D. Foata, "Rearrangement of words," in: M. Lothaire, *Combinatorics on Words*, Encyclopedia of Mathematics and its Applications, vol. 17, Addison-Wesley, Reading, Mass., 1983.

[Françon74] J. Françon, "Preuves combinatoires des identités d'Abel," *Discrete Mathematics*, **8**(1974), 331-343.

[Gertsbakh77] I. B. Gertsbakh, "Epidemic processes on a random graph: Some preliminary results," *Journal of Applied Probability*, **14**(1977), 427–438.

[GK68] B. V. Gnedenko and A. N. Kolmogorov, *Limit Distributions for Sums of Independent Random Variables*, Second edition, Addison-Wesley, Cambridge, Mass., 1968.

[GS83] R. Gold and J. Sattler, "Modifikationen des Pollard-Algoritmus," *Computing*, **30**(1983), 77-89.

[Harris60] B. Harris, "Probability distributions related to random mappings," *Annals of Mathematical Statistics*, **31**(1960), 1045–1062.

[Hellman80] M. E. Hellman, "A cryptanalytic time-memory tradeoff," *IEEE Transactions on Information Theory*, **IT-26**(1980), 401–406.

[HLP59] G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities*, Second edition, Cambridge University Press, 1959.

[IM65] G. I. Ivchenko and Yu. I. Medvedev, "Asymptotic representations of finite differences of a power function at an arbitrary point," *Theory of Probability and its Applications*, **10**(1965), 139–144.

[Jordan47] C. Jordan, *Calculus of Finite Difference*, Chelsea, New York, 1947.

[Karnin83] E. D. Karnin, "Probabilistic and computational methods in cryptography," Technical report, Information Systems Laboratory, Stanford University, 1983.

[Katz55] L. Katz, "Probability of indecomposability of a random mapping function," *Annals of Mathematical Statistics*, 26(1955), 512–517.

[Knuth73a] D. E. Knuth, *The Art of Computer Programming*, Vol. 1, Second edition, Addison-Wesley, Reading, Mass., 1973.

[Knuth73b] D. E. Knuth, *The Art of Computer Programming*, Vol. 3, Addison-Wesley, Reading, Mass., 1973.

[Knuth81] D. E. Knuth, *The Art of Computer Programming*, Vol. 2, Second edition, Addison-Wesley, Reading, Mass., 1981.

[Knuth85] D. E. Knuth, "The analysis of optimum caching," *Journal of Algorithms*, 6(1985), 181–199.

[Koutras82] M. Koutras, "Non-central Stirling numbers and some applications," *Discrete Mathematics*, 42(1982), 73–89.

[KR75] D. E. Knuth and G. S. Rao, "Activity in an interleaved memory," *IEEE Transactions on Computers*, C-24(1975), 943–944.

[Kruskal54] M. D. Kruskal, "The expected number of components under a random mapping function," *American Mathematical Monthly*, 61(1954), 392–397.

[KS78] D. E. Knuth and A. Schönhage, "The expected linearity of a simple equivalence algorithm," *Theoretical Computer Science*, 6(1978), 281–315.

[Moon71] J. W. Moon, *Counting Labelled Trees*, Canadian Mathematical Monographs, 1971.

[Nielsen23] N. Nielsen, *Traité Élémentaire des Nombres de Bernoulli*, Gauthier-Villars, Paris, 1923.

[Patil60] G. P. Patil, "On the evaluations of the negative binomial distribution with examples," *Technometrics*, 2(1960), 501–505.

[Pearson33] K. Pearson, "On the applications of the double Bessel function $K_{r_1,r_2}(x)$ to statistical problems," *Biometrika*, 25(1933), 158–178.

[Petrov75] V. V. Petrov, *Sums of Independent Random Variables*, Springer-Verlag, New York/Heidelberg/Berlin, 1975.

[Pittel83] B. Pittel, "On distributions related to the transitive closures of random finite mappings," *Annals of Probability*, **11**(1983), 428–441.

[Pollard75] J. M. Pollard, "A Monte Carlo method for factorization," *BIT*, **15**(1975), 331–334.

[Pollard82] J. M. Pollard, Personal communication to D. E. Knuth, 1982.

[Pollard83] J. M. Pollard, Personal communication, 1983.

[Pomerance82] C. Pomerance, "Analysis and comparison of some integer factoring algorithms," in R. Tijdeman, and H. W. Lenstra (eds.), *Computational Methods in Number Theory*, Math. Centrum Tracts, Number 154 (Part I), and Number 155 (Part II), Amsterdam, 1982.

[Ramanujan12] S. Ramanujan, "Questions for solution, number 294," *Journal of the Indian Mathematical Society*, **4**(1912), 151–152.

[Riordan58] J. Riordan, *An Introduction to Combinatorial Analysis*, Wiley, New York, 1958.

[Riordan68] J. Riordan, *Combinatorial Identities*, Wiley, New York, 1968.

[Riordan69] J. Riordan, "Abel identities and inverse relations," in: R.C. Bose and T.A. Dowling, eds., *Combinatorial Mathematics and its Application*, Univ. of North Carolina Press, Chapel Hill, 1969.

[SL84] C. P. Schnorr and H. W. Lenstra, Jr., "A Monte Carlo factoring algorithm with linear storage," *Mathematics of Computation*, **43**(1984), 289–311.

[SSY82] R. Sedgewick, T. G. Szymansky, and A. C. Yao, "The complexity of finding cycles in periodic functions," *SIAM Journal on Computing*, **11**(1982), 376–390.

[Stanford81] Stanford Computer Science Department, Qualifying examination in the analysis of algorithms, April 1981.

[Stepanov69] V. E. Stepanov, "Limit distributions of certain characteristics of random mappings," *Theory of Probability and its Applications*, **14**(1969), 612–626.

[Zave76] D. A. Zave, "A series expansion involving harmonic numbers," *Information Processing Letters*, **5**(1976), 75–77.

Appendix A

A bibliography of random mappings

This appendix is not an exhaustive list of the literature on random mappings, but might constitute a starting point for an exhaustive search. In particular, results on random graphs that are not specific to random mappings were omitted. Papers already mentioned in the "References" section are marked with an asterisk.

[1] A. S. Ambrosimov, "The distribution of the number of nonappearing lengths of cycles in a random mapping," *Matematicheskie Zametki*, **23**(1978), 895–898.

[2*] S. Berg, "On snowball sampling, random mappings and related problems," *Journal of Applied Probability*, **18**(1981), 283–290.

[3] Y. D. Burtin, "On a simple formula for random mappings and its applications," *Journal of Applied Probability*, **17**(1980), 403–414.

[4*] L. Comtet, *Advanced Combinatorics*, Reidel, Dordrecht/Boston, 1974.

[5] J. E. Folkert, "The distribution of the number of components of a random mapping function," unpublished Ph. D. dissertation, Michigan State University, 1955.

[6] A. E. Gelfand, "On the cyclic behavior of random transformations on a finite set," Technical report no. 305, Department of Statistics, Stanford University, 1981.

[7*] I. B. Gertsbakh, "Epidemic processes on a random graph: some preliminary results," *Journal of Applied Probability*, **14**(1977), 427–438.

[8] A. A. Grusho, "Random mappings with bounded multiplicity," *Theory of Probability and its Applications*, **17**(1972), 416–425.

[9*] B. Harris, "Probability distributions related to random mappings," *Annals of Mathematical Statistics*, **31**(1960), 1045–1062.

[10*] E. D. Karnin, "Probabilistic and computational methods in cryptography," Technical report, Information Systems Laboratory, Stanford University, 1983.

[11*] D. E. Knuth, *The Art of Computer Programming*, Vol. 2, Second edition, Addison-Wesley, Reading, Mass., 1981.

[12*] M. D. Kruskal, "The expected number of components under a random mapping function," *American Mathematical Monthly*, **61**(1954), 392–397.

[13] N. Metropolis, S. Ulam, "A property of randomness of an arithmetical function," *American Mathematical Monthly*, **60**(1953), 252–253.

[14*] J. W. Moon, *Counting Labelled Trees*, Canadian Mathematical Monographs, 1971.

[15] L. R. Mutafchiev, "Limit properties of components of random mappings," *Comptes Rendus de l'Academie Bulgare des Sciences*, **31**(1978), 1257–1260.

[16] Yu. L. Pavlov, "Limit theorems for a characteristic of a random mapping," *Theory of Probability and its Applications*, **26**(1982), 829–834.

[17*] B. Pittel, "On distributions related to the transitive closures of random finite mappings," *Annals of Probability*, **11**(1983), 428–441.

[18] A. Rapaport, "Cycle distribution in random nets," *Bulletin of Mathematical Biophysics*, **10**(1948), 145–157.

[19] S. M. Ross, "A random graph," *Journal of Applied Probability*, **18**(1981), 309–315.

[20] H. Rubin and R. Sitgreaves, "Probability distributions related to random transformations on a finite set," Technical report no. 19, Applied Mathematics and Statistics Laboratory, Stanford University, 1954.

[21*] R. Sedgewick, T. G. Szymansky, and A. C. Yao, "The complexity of finding cycles in periodic functions," *SIAM Journal on Computing*, **11**(1982), 376–390.

[22*] V. E. Stepanov, "Limit distributions of certain characteristics of random mappings," *Theory of Probability and its Applications*, **14**(1969), 612–626.

[23] V. E. Stepanov, "Random mappings with a single attracting centre," *Theory of Probability and its Applications*, **16**(1971), 155–161.

Appendix B

A strange example

This an example of a permutation invariant weight for which the correlation of λ and μ is positive. Consider the following probability weight on $F[n]$:

$$w(f) = \begin{cases} \epsilon, & \text{if } C_0(f) \text{ is a permutation of } 1,1,\ldots,1,2,2,3; \\ 1-n(n-1)(n-2)\epsilon/2, & \text{if } C_0(f) = 1,1,\ldots,1; \\ 0, & \text{otherwise.} \end{cases}$$

Clearly, this weight is permutation invariant. The probability generating function for the number of cyclic elements is

$$C(z) = \left(1-(3n^2-13n+14)\epsilon\right)z + (3n^2-19n+32)z^2 + 6(n-3)z^3.$$

Using the relevant equations from Chapter 1, it can be shown that

$$\widehat{\mathrm{cov}}(\lambda,\mu) = \epsilon\left(\frac{n}{2} - \frac{29}{2} + \frac{38}{n}\right) - \epsilon^2\left(\frac{9n^4}{4} + \frac{39n^3}{2} - \frac{157n^2}{4} - 37n + 103 + \frac{52}{n}\right).$$

Therefore if, say, $n = 100$ and $\epsilon = 10^{-7}$, then the covariance is positive. (Namely, it is equal to $0.000001529\ldots$.)

END
FILMED

DATE:

5-8-96

NTIS